

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 17, 2002, 15:35:20 : Search time 3293 Seconds  
(without alignments)  
10755.577 Million cell updates/sec

Title: US-09-720-384A-3

Perfect score: 1217

Sequence: 1 ggcgcgtttcattcatca.....aaaaaaaaaaaaaaaaaaaa 1217

Scoring table: OLIGO NUC

Gapop'60.0, Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size: 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.mu.\*

20: em.om.\*

21: em.or.\*

22: em.ov.\*

23: em.pat.\*

24: em.ph.\*

25: em.pl.\*

26: em.ro.\*

27: em.sts.\*

28: em.un.\*

29: em.vi.\*

30: em.htg.hum.\*

31: em.htg.in.\*

32: em.htg.other.\*

33: em.htg.mu.\*

34: em.htg.pln.\*

35: em.htg.rod.\*

36: em.htg.man.\*

37: em.htg.vrt.\*

38: em.sy.\*

39: em.htgo.hum.\*

40: em.htgo.mu.\*

41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	39	3.2	154330	2	AC104429	AC104429 Oryza sat
2	35	2.9	931	9	BC010286	BC010286 Homo sapi
3	35	2.9	1069	3	AY061385	AY061385 Drosophila
4	35	2.9	113880	3	PFMAL3P4	AL008970 Plasmodium
5	35	2.9	116696	3	PFMAL3P3	Z98547 Plasmodium
6	35	2.9	149710	2	AC022563	AC022563 Homo sapi
7	35	2.9	203610	2	AC102579	AC102579 Mus muscu
8	34	2.8	57	6	AX074093	AX074093 Sequence
9	34	2.8	57	6	AX078311	AX078311 Sequence
10	34	2.8	57	6	AX397797	AX397797 Sequence
11	34	2.8	540	3	AY113210	AY113210 Drosophila
12	34	2.8	853	3	AY119449	AY119449 Drosophila
13	34	2.8	1551	3	AY060916	AY060916 Drosophila
14	34	2.8	1627	9	BC036471	BC036471 Homo sapi
15	34	2.8	1771	3	AY070713	AY070713 Drosophila
16	34	2.8	1852	3	AF132556	AF132556 Drosophila
17	34	2.8	1908	6	AX045627	AX045627 Sequence
18	34	2.8	3165	3	AY089433	AY089433 Drosophila
19	34	2.8	7128	6	AX346460	AX346460 Sequence
20	34	2.8	14566	10	AF433638	AF433638 Mus muscu
21	34	2.8	40324	6	AX458633	AX458633 Sequence
22	34	2.8	60599	2	AC129924	AC129924 Homo sapi
23	34	2.8	62964	2	AC104959	AC104959 Homo sapi
24	34	2.8	63988	2	AC103719	AC103719 Homo sapi
25	34	2.8	128379	9	HSJ773A18	AL049557 Human DNA
26	34	2.8	137921	9	AC025048	AC025048 Homo sapi
27	34	2.8	140233	2	AC009422	AC009422 Homo sapi
28	34	2.8	153023	2	AC021462	AC021462 Homo sapi
29	34	2.8	155689	2	AC097248	AC097248 Rattus no
30	34	2.8	215126	2	AC083916	AC083916 Mus muscu
31	34	2.8	231209	2	AL807831	AL807831 Mus muscu
32	33	2.7	307	11	G73625	G73625 RZ742F etio
33	33	2.7	480	10	BC028514	BC028514 Mus muscu
34	33	2.7	501	10	BC024343	BC024343 Mus muscu
35	33	2.7	531	8	AF064734	AF064734 Dianthus
36	33	2.7	535	8	SSI243446	AJ243446 Spermatoz
37	33	2.7	550	3	TGSLACS1	X06346 T. gambiense
38	33	2.7	583	9	BC000838	BC000838 Homo sapi
39	33	2.7	608	6	AX186574	AX186574 Sequence
40	33	2.7	617	8	AF323605	AF323605 Manihot e
41	33	2.7	638	9	BC007043	BC007043 Homo sapi
42	33	2.7	641	8	AF266463	AF266463 Manihot e
43	33	2.7	762	8	AT092084	U92084 Arabidopsis
44	33	2.7	773	8	AF380357	AF380357 Porteresi
45	33	2.7	786	9	BC003564	BC003564 Homo sapi

ALIGNMENTS

RESULT 1  
AC104429  
LOCUS  
DEFINITION  
Oryza sativa chromosome 3 clone OSJNB0006P09, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 8 ordered pieces.  
ACCESSION  
AC104429  
VERSION  
AC104429.1 GI:17488564  
KEYWORDS  
HTG; HTGS PHASE2.  
SOURCE  
Oryza sativa  
ORGANISM  
Oryza sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.  
REFERENCE  
1 (bases 1 to 154330)  
AUTHORS  
Wing, R.A., Yu, Y., Soderlund, C., Chen, M., Kim, H.-R., Rambo, T.,

TITLE Sasaki, C., Henry, D., Oates, R. and Simmons, J.  
JOURNAL Rice Genomic Sequence  
REFERENCE 2 (bases 1 to 154330)  
AUTHORS Wang, R. A., Yu, Y., Soderlund, C., Chen, M., Kim, H.-R., Rambo, T., Sasaki, C., Henry, D., Oates, R. and Simmons, J.  
TITLE Direct Submission  
JOURNAL Submitted (11-DEC-2001) Clemson University Genomics Institute, Clemson University, 100 Jordan Hall, Clemson, SC 29634, USA  
COMMENT \* NOTE: This is a 'working draft' sequence. It currently consists of 8 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.  
\* This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.  
\* 1 1981: contig of 1981 bp in length  
\* 1982 32672: contig of 30691 bp in length  
\* 32673 87321: contig of unknown length  
\* 87322 94650: contig of 54649 bp in length  
\* 94651 96734: contig of unknown length  
\* 96735 96734: contig of 2084 bp in length  
\* 96735 97792: gap of unknown length  
\* 97793 139678: gap of 1058 bp in length  
\* 139679 154330: gap of unknown length  
\* 154330: contig of 41886 bp in length  
\* 139679 154330: contig of 14652 bp in length.  
LOCATION/Qualifiers  
1. 154330  
/organism="Oryza sativa"  
/db\_xref="taxon:4530"  
/chromosome="3"  
/clone="OSUNB0006P09"  
/clone="C3J3401 C 3J3796 g 43714 t 10 others  
BASE COUNT 43409 a 33401 c 33796 g 43714 t 10 others  
ORIGIN  
Query Match 3.2%; Score 39; DB 2; Length 154330;  
Best Local Similarity 100.0%; Pred. No. 1.7e-10;  
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 647 AAGCGAAGACCCGCGAAGAAATATACGAAGACTGCT 685  
Db 121578 AAGCGAAGACCCGCGAAGAAATATACGAAGACTGCT 121516  
RESULT 2  
LOCUS BC010286 931 bp mRNA linear PRI 29-OCT-2001  
DEFINITION Homo sapiens, ribosomal protein S4, y-linked, clone MGC:5070  
ACCESSION BC010286  
VERSION BC010286.1 GI:16307476  
KEYWORDS MGC.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS 1 (bases 1 to 931)  
TITLE Direct Submission  
JOURNAL Submitted (05-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNLI)  
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [villalobosbcm.tmc.edu](mailto:villalobosbcm.tmc.edu)  
Villalobos, D.K., Luna, R.A., Hale, S.M., Huliyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNLI at: <http://image.llnl.gov>  
Series: IRAC Plate: 3 Row: n Column: 21  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 7959920.  
LOCATION/Qualifiers  
1. 931  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/db\_xref="LocustID:6192"  
/db\_xref="taxon:9606"  
/clone="MGC:5070 IMAGE:3048929"  
/clone\_type="Placenta, choriocarcinoma"  
/clone\_id="NIH MGC\_10"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
13. 804  
/codon\_start=1  
/product="ribosomal protein S4, y-linked"  
/protein\_id="AAH10286.1"  
/db\_xref="GI:16307477"  
/translation="MARCPKHLKRVAAKPHMMLDKLTGVPAFPTSTGPHKLRECLPLIVFLRNRLKYALTGDEVEKICMQRFIKIDGKRVVDVTPAFPMFVVISIEKTSHEFLVYDFKRFPAVHRTVEBAKYLCKVKRKTIVGKGIPLVTHAPRTIRYDPDVIVKNTVQIDLGKXINFIKFDGTLNLCMVYGGANLGRVGTITNERHSGSFVDVHVKDANGNSFATLSNIFPIGNKPKWISLPRGKIRLTVAERBRKRLATQSSG"  
BASE COUNT 302 a 183 c 224 g 222 t  
ORIGIN  
Query Match 2.9%; Score 35; DB 9; Length 931;  
Best Local Similarity 100.0%; Pred. No. 2.7e-08;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0.  
Oy 1183 TCCTTTAAAAA  
Db 856 TCCTTTAAAAA  
RESULT 3  
LOCUS AY061385 1069 bp mRNA linear INV 08-NOV-2001  
DEFINITION Drosophila melanogaster LD30889 full length cDNA.  
ACCESSION AY061385  
VERSION AY061385.1 GI:16769427  
KEYWORDS FLI cDNA.  
SOURCE Drosophila melanogaster.  
ORGANISM Drosophila melanogaster.  
REFERENCE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
AUTHORS 1 (bases 1 to 1069)  
TITLE Direct Submission  
JOURNAL Submitted (30-OCT-2001) Berkeley Drosophila Genome Project, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, USA  
COMMENT Sequence submitted by: Berkeley Drosophila Genome Project

## Laurence Berkeley National Laboratory

Berkeley, CA 94720  
This clone was sequenced as part of a high-throughput process to sequence clones from Drosophila Gene Collection 1 (Rubin et al., Science 2000). The sequence has been subjected to integrity checks for sequence accuracy, presence of a polyA tail and contiguity within 100 kb in the genome. Thus we believe the sequence to reflect accurately this particular cDNA clone. However, there are artifacts associated with the generation of cDNA clones that may have not been detected in our initial analyses such as internal priming, priming from contaminating genomic DNA, retained introns due to reverse transcription of unspliced precursor RNAs, and reverse transcriptase errors that result in single base changes. For further information about this sequence, including its location and relationship to other sequences, please visit our Web site (<http://fruitfly.berkeley.edu>) or send email to [cdna@fruitfly.berkeley.edu](mailto:cdna@fruitfly.berkeley.edu)

FEATURES  
source  
1..1069  
/organism="Drosophila melanogaster"  
/strain="y; cn bw sp"  
/db\_xref="taxon:7227"  
misc\_feature  
1..1031  
/notes="blastn alignment with AE003431"  
84..233  
/notes="Longest ORF"  
/codon\_start=1  
/product="LD30889p"  
/protein\_id="AAL28933.1"  
/db\_xref="GI:16769438"  
/translation="MQEIAGLIISCGLKFFFFFLNSLSNHTMHKLGITFGISGE  
KGTPK"  
BASE COUNT 325 a 243 c 197 g 304 t  
ORIGIN  
Query Match 2.9%; Score 35; DB 3; Length 1069;  
Best Local Similarity 100.0%; Pred. No. 2.7e-08;  
Matches - 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1183 TCGTTTAAAAA...AAAAAAAAAAAAAAAAAAAA 1217  
|||  
Db 1026 TCGTTTAAAAA...AAAAAAAAAAAAAAAAAAAA 1060  
|||

RESULT 4  
PFMAL3P4  
LOCUS  
DEFINITION  
Plasmodium falciparum MAL3P4, complete sequence.  
ACCESSION  
AL008970 AL339179  
VERSION  
AL008970.3 GI:7672212  
KEYWORDS  
HTG.  
SOURCE  
Plasmodium falciparum 3D7.  
ORGANISM  
Plasmodium falciparum 3D7.  
REFERENCE  
1 (bases 1 to 113880)  
Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.

AUTHORS  
Bowman, S., Lawson, D., Basham, D., Brown, D., Chillingworth, T., Churcher, C.M., Craig, A., Davies, R.M., Devlin, K., Feltwell, T., Gentles, S., Gwilliam, R., Hamlin, N., Harris, D., Holroyd, S., Hornsby, T., Horrocks, P., Jagels, K., Jassal, B., Kves, S., McLean, J., Moule, S., Mungall, K., Murphy, L., Oliver, K., Quail, M.A., Rajandream, M.-A., Rutter, S., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Whitehead, S., Woodward, J.R., Newbold, C. and Barrell, B.G.  
The complete nucleotide sequence of chromosome 3 of Plasmodium falciparum  
Nature 400 (6744), 532-538 (1999)  
JOURNAL  
MEDLINE  
99376085  
PUBMED  
10448855  
REFERENCE  
2 (bases 1 to 113880)  
AUTHORS  
Hamlin, N., Bowman, S., Churcher, C., Lawson, D., Quail, M. and Barrell, B.  
JOURNAL  
Unpublished  
REFERENCE  
3 (bases 1 to 113880)

## AUTHORS

Lawson, D., Bowman, S. and Barrell, B.  
Direct Submission  
Submitted (18-NOV-1998) P.falciparum Genome Sequencing Consortium,  
The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge  
CB10 1SA, UK

COMMENT  
On Apr 30, 2000 this sequence version replaced gi:4493924.  
For more information about this sequence or the Malaria Project,  
see [http://www.sanger.ac.uk/Projects/P\\_falciparum](http://www.sanger.ac.uk/Projects/P_falciparum).

## FEATURES

## source

1..113880  
/organism="Plasmodium falciparum 3D7"  
/strain="3D7"  
/db\_xref="taxon:36329"

/chromosome="3"  
/clone="MAL3P4"

1..40

/note="original nominal overlap with MAL3P3"

## misc\_feature

41..2000

/note="sequence duplicated in MAL3P3 to include complete CDS"

## gene

2364..6248  
/gene="PFC0435w, MAL3P4.26"

2364..6248  
/gene="PFC0435w, MAL3P4.26"

/note="PFC0435w (MAL3P4.26), Hypothetical protein, len: 1294 aa, predicted using hexExon"

/codon\_start=1

/product="hypothetical protein, PFC0435w"

/protein\_id="CAAJ5616.1"

/db\_xref="GI:3764023"

/translation="MKVGIIFFCLFFVVLGACNVNKRIFKIKKRTFIILNBPV  
DUSFENLFTLLDLDVLDKLYLDESLLNENLVSSIFRLLDVDTYKNIKENEDDN  
KNIRYIFLGTSFISIRHPANFEYFLKANKIYNGIYKGNVDIRGILYEKNIIEEK  
KLEKQAKIKDKNNNNNNNSKFSKGDENFNKNDLYNPSDKLYNNDDIDVHE  
LLEEIITKEKRFLLDDDDNDNDKYLKTDENVKYGFGIFGDFNDIIPSVIHKENLRV  
DKNFIPLNSGIIILDTLTKNIYEVSNILSNKQDSIHDIYEVTKYIKENLRV  
RUTHSENCLNEONIHLLDNPNNFIYKYQVILNFKDYNKTEKQYKIGHEVIR  
RUEETSSGENLNRTNKHNDNNNNNSYEDAEILLSYNVEYFVSTCMCYISIR  
SKHSLMDYDKYMINLENDIKLKHVETBEIHPNSIEEYKMLARINUKYDITLLEE  
HENLVTHKNLIGIKTSINTERERIPIKNTYDNKENTQIIFNFYDNKLEKNTFFG  
FYNSLQNALENDNIDLIYMSDKSKYDNLNFYSKTSREGCEKLGKMIYIY  
EYVWKNSEKYPFIADDDTFVNVNKLIDVTLNLTCSHKKYMDYDKYIKSYDFVYKE  
NEALFLQNFPRKTLFLYSLKDTFAKTIQTLKKYDYVPCQGGIISKKHNNDSDD  
HDHVGKQNNDSNHQDIENQVNVNNNNNNNNKAKSIPIYLGERSYNTFTSN  
NRYFYDLTGAGILINDEAKRIYEKECTCPTSNSSMDMIFGWAKELGILAINF  
EGYFQNSPLDYNNKYINTLPIYTHRLNKRITTKESRDMYFNLYNVRNDEKQNDI  
YVDYLDRHKMIDNVPHYFYVMYDEKNVTVTKIEHNADMSKKNSKNPKLNT  
QGDKNVDNNDENVKGNVKGDNVKGDNVKGDNVKGDNVKGDNVKGDNVKGDNV  
NKGDDNNYVNDNNDIINNVSVDVDDUMERNKKGCTKEKKDKDKNHNKEKATDV  
KKSFTVNNIDNEDDTTKYIKNKEIKYNRMQESGKYQLFDINKDKFKKEIGHPYFQ  
KIKKNKAKKEKEMKQKKQKDYTNFTFTSNQGNFQGNQGNFQGNQGNFQGNQGN  
DQRPBIEADINPMDEYEWENLSNFDGDEPYDEYDDYDDVNTINADKLKINDQNK  
HYEQIKDIAQFPVFNQDQNSNTFDFDDEL"

6245..6250  
/note="potential splice donor sequence at end of PFC0435w  
may suggest splicing"

complement(7159..7164)

/note="potential splice donor sequence at end of PFC0440c  
may suggest splicing"

complement(7161..13970)

/gene="PFC0440c, MAL3P4.25"

complement(7161..13970)

/note="PFC0440c (MAL3P4.25), putative helicase, len: 2270  
aa; similarity to helicases. C.elegans helicase  
(WP:FS2B5.3) BLAST Score: 290, sum P(4) = 5.9e-31; 34% in  
199 aa overlap, predicted using hexExon, contains match to  
PFC0271 helicase\_C, Helicases conserved C-terminal domain  
Score 25.70"

/codon\_start=1

/product="putative helicase"

/protein\_id="CAAJ5615.1"

## misc\_feature

7159..7164

/note="potential splice donor sequence at end of PFC0440c  
may suggest splicing"

complement(7161..13970)

/gene="PFC0440c, MAL3P4.25"

complement(7161..13970)

/note="PFC0440c (MAL3P4.25), putative helicase, len: 2270  
aa; similarity to helicases. C.elegans helicase  
(WP:FS2B5.3) BLAST Score: 290, sum P(4) = 5.9e-31; 34% in  
199 aa overlap, predicted using hexExon, contains match to  
PFC0271 helicase\_C, Helicases conserved C-terminal domain  
Score 25.70"

/codon\_start=1

/product="putative helicase"

/protein\_id="CAAJ5615.1"

## misc\_feature

7159..7164

/note="potential splice donor sequence at end of PFC0440c  
may suggest splicing"

complement(7161..13970)

/gene="PFC0440c, MAL3P4.25"

complement(7161..13970)

/note="PFC0440c (MAL3P4.25), putative helicase, len: 2270  
aa; similarity to helicases. C.elegans helicase  
(WP:FS2B5.3) BLAST Score: 290, sum P(4) = 5.9e-31; 34% in  
199 aa overlap, predicted using hexExon, contains match to  
PFC0271 helicase\_C, Helicases conserved C-terminal domain  
Score 25.70"

/codon\_start=1

/product="putative helicase"

/protein\_id="CAAJ5615.1"

## misc\_feature

7159..7164

/note="potential splice donor sequence at end of PFC0440c  
may suggest splicing"

complement(7161..13970)

/gene="PFC0440c, MAL3P4.25"

complement(7161..13970)

/note="PFC0440c (MAL3P4.25), putative helicase, len: 2270  
aa; similarity to helicases. C.elegans helicase  
(WP:FS2B5.3) BLAST Score: 290, sum P(4) = 5.9e-31; 34% in  
199 aa overlap, predicted using hexExon, contains match to  
PFC0271 helicase\_C, Helicases conserved C-terminal domain  
Score 25.70"

/codon\_start=1

/product="putative helicase"

/protein\_id="CAAJ5615.1"

/db\_xref="GI:3764022"  
/db\_xref="SPTREMBL:O77360"  
/translation="MANTFKINKKRRRTYVESYNIINIAKKKAYMCDNRKNDNINII  
EKYFNSSSKIKTGKNNYKNDIKSYDMCEERSMDDEENIYNNYKGRKNNYSK  
NNYSNNIINHNDNRNTYGGD.DGRKOYLSNKHYSKNDNENYLNKKTNPPIPIKYN  
NNNNNNNNNNNDNRKNDINNTYKPYDKNEKSKNTHSYGNPNNEKTPNDPKOYM  
SNESHNTQDGHPTIYNHNNNDGRIKRKQDISLSSTGYRKISNDLSKENTSYNY  
NNSGGAARHVAHQNNQNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
QNDMDKNYYODIHKNNMKSSTKQDTA1SFNGKVFSPKQKMEKSKFFKSNRNE  
SINPGMNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
MPIYNAENCNLKKM1YKDKDEKEKIKILTDEPNVQSNIEKLSYKSRNEIEM  
TEKDVTFINGETSGSKTCVCKPLEENIRENNKINI1VTEPRRIADIALSKISLST  
TNEKLGOKIYGRISGESLYDSEKTYITITGYFLPLFHGMKAKTHYI1DEID  
RSILDTVLITFLYKLNHKKOQDEOMFKLIMSATMOQSNFYGFHPMKIKMSIPIGT  
KISIDTPTFYEDINATYRGSNNIDKDEKINDKINDKINDKINDKINDKINDKIND  
DKSACDCKSCDGNRFKESYIDPTLRCKCKINLSKQSEMLYKIKSYEDKIH  
FNNNNKYCNDKRDLDVEITPAVFSNLSNLCLELVNLCKGDSYITLSGQDI  
TDMYHQLSMIINNNDNPSNNININNGNDANDANDVNNHNNHNNHNNHNNHNNH  
HRDICKIHMHSCLYDNTJHKLHNDTINI1FLSNIAESSITI1PVRNLVIDFCIO  
KNIEYNDKKKAHILVAKMINKSMEORCGRCGICIRMI1SKQFNILRDHKISE  
IYTHSLHLYLYLKSMSYVNLGINKNEVHCNTNVAVKEVDQKNGIKNEISH  
SHAENKKTJYDYL5WITEKSEKIKSTRELEKAVIKIKDKLVISIGQIMIEF  
NLSTNICRLILGVLDVPTFTIILIGILNTDIPENINLSSKRIYSVAISLEICV  
OKAFPCNTYSPRIMRVNFFLEMLCYVLVQGLKKENFHRKELKTYNNMTCINMK  
RNNVNSKILICVINSYDNLCKMKMLKNSNAKSCYVLLVLLGSGDINYNMNS  
NNNDNSGETINADGTPNTHMNSINGNNVFNVTCTLPDYSQNTYLFELFSLF  
TPLEFHGSPMLPLRLNLEGKSKKLMANVLFMI1DKLDVKNCITFSQVYKTI1CDKGD  
RALCIMCPYLSFDIYVSK1YVYFHNNNNNHNNN1LKGD1FSYKNTI1PDKNAD  
MSVNSYLNLTQEGHDTGITSFGVSPLMNNNNNNNNY1HNNVY1PCGINNNYNN  
VNSQNNNNMMKIMSNYSSESFLNTGSGVFDI1EIKIRKLKYKNNNEVSIPHN  
NNI1NLTOPNAIVPYNSKDEINNAIS1CNII1NENSGKI1FPM1FENKSKCY  
SEYFNHKEFLKPKRHLFLYKMPFLNTDYYKIKKQKEHMPNTNALNNNSLKXNS  
KNQIHRNINDVSLNDQEDQYNGYVYDDYDDDDNDNDNNNNNNNNNDNN  
NNNDYDYSVYKSNNTTNNNDPQYEGNNSGNNNNNNNNNNNNNNNNNNNNNN  
LGFISKCPFRPTKRNITNQTAD1YACSSIDYSTNTYNNVYVLEPNKYELTF  
FLSIPYHNVIT1QTRNL1YSDI1STIKIPSKELILKMKMKKAKKTNTTYTDTN  
STINNPT1KHDILRVNLYRVMSKLLSLTSPNQEL1EONTOPLKSEKSKASY  
INKGPNNDNRVANDKVNQNDKVNNDQNDVNDQNDQNDHDFDEDEYFSD  
IMSDIENNNILKENLYLKOIVIMTEENNDY1NYSKNDMDIDEHFNLYDEDE  
HFILNKEKYNKGEILMNTYDLYENS1KENNDYNNNNNNNNNNNNNNNNNNNN  
KEXKLNLEAVKSSYLNNSFVQONTMNAVERTSEDFLPQPINISPIKEANYKLR  
SMYNEVAPLDFKFRTRNQV"  
CDS  
/gene="PFC0441C"  
/note="variable sequence, 9 or 10 Ts"  
/complement(join(16039..16667,16813..16891))  
/gene="PFC0441C"  
/complement(join(16039..16667,16813..16891))  
/note="PFC0441C, hypothetical protein, len: 235 aa,  
revised: new gene prediction"  
variation  
/codon\_start=1  
/product="hypothetical protein, PFC0441C"  
/protein\_id="CAB89483.1"  
/db\_xref="GI:7672213"  
/translation="MDINNIMEEVOAIVLPMMFNASKLVLSYVNAFENNEYD  
NDWLTLRFLCYPCGYDKIIRKILCVYINNVNMMYMSLINSLSLNDIKYV  
IYIELIKNGOYKLLKNNINNNINNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
SIVYKMSSEYLNLTQDTEOFLNENKTKIKM1NHGSKQEOICVGNMETVQNNKNIN  
TYTFEDNIGSYITKLNH"  
misc\_feature  
/complement(16668..16677)  
/gene="PFC0441C"  
/note="predicted splice acceptor sequence for exon 2 of  
PFC0441C"  
misc\_feature  
/complement(16807..16812)  
/gene="PFC0441C"  
/note="predicted splice donor sequence for exon 1 of  
PFC0441C"  
gene  
join(18192..18242,18754..19140)  
/gene="PFC0445W, MAL3P4.24"  
/db\_xref="SPTREMBL:O77359"

CDS  
join(18192..18242,18754..19140)  
/gene="PFC0445W, MAL3P4.24"  
/note="PFC0445W, MAL3P4.24, conserved hypothetical  
protein, len: 145 aa, revised: extended upstream and  
added exon, GC plots suggest more, similarity: eg to human  
cDNA AAF29135 HSPC172 (219 aa), fasta scores: opt: 257,  
E(): 6.9e-11 (40.0% identity in 115 aa overlap)"  
/codon\_start=1  
/product="hypothetical protein, PFC0445W"  
/protein\_id="CAA15613.2"  
/db\_xref="SPTREMBL:O77358"  
/db\_xref="SPTREMBL:O77358"  
/translation="MYSLVNNHOGTAVYQKHFSEIKLNSNEIRLASMIGISTIS  
EKINVYLYKKTNI1FOSLEKGIETIEGDFKIOCYDTLTGIRK1FVHKDNLIEWN  
TYLKRVELYSDIILKNNPVDIMPIRSVAVFNEQIKLPSNLS"  
19528..19857  
/gene="PFC0450W, MAL3P4.23"  
19528..19857  
/gene="PFC0450W, MAL3P4.23"  
/note="PFC0450W, MAL3P4.24, Hypothetical protein, len:  
109 aa, contains possible signal sequence, predicted using  
hexon"  
/codon\_start=1  
/product="hypothetical protein, PFC0450W"  
/protein\_id="CAA15614.1"  
/db\_xref="GI:3764021"  
/db\_xref="SPTREMBL:O77359"  
Query Match 2.9%; Score 35; DB 3; Length 113880;  
Best Local Similarity 100.0%; Pred. No. 3.3e-08;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1183 TCGTTAAAAA.....AAAAAAAAA 1217  
DB 1690 TCGTTAAAAA.....AAAAAAAAA 1724  
RESULT 5  
LOCUS PFMAL3P3 116696 bp DNA linear INV 27-APR-2000  
DEFINITION Plasmodium falciiparum MAL3P3, complete sequence.  
ACCESSION Z985447 AL008977 AL010164 AL139179 Z98548 Z98549 Z98550  
VERSION Z98547.2 GI:1538394  
KEYWORDS HTG: 60S Acidic ribosomal protein P2; asparagine synthetase;  
calcium-dependent ribose kinase; DNA polymerase delta; dual  
specificity protein phosphatase; guanine nucleotide-binding  
protein; PDZ domain; rat BRAIN-like protein; secine/threonine  
protein kinases; T-complex protein eta; T08A11.2-like protein; WD  
domain; Y48E1C.2-like protein.  
SOURCE Plasmodium falciiparum 3D7.  
ORGNISM Plasmodium falciiparum 3D7.  
REFERENCE 1 (bases 1 to 116696)  
AUTHORS Bowman,S., Lawson,D., Baaham,D., Brown,D., Chillingworth,T.,  
Churcher,C.M., Craig,A., Davies,R.M., Devlin,K., Feltham,T.,  
Genies,S., Gwilliam,R., Hamlin,N., Harris,D., Holtz,S.,  
Horsby,T., Horrocks,P., Jagers,K., Jaseal,B., Kyes,S., McLean,J.,  
Mout,R., Mungall,K., Murphy,L., Oliver,K., Quail,M.A.,  
Rajandream,M.-A., Rutter,S., Skelton,J., Squares,S.,  
Sulston,J.E., Whitehead,S., Woodward,J.R., Newbold,C. and  
Barrell,B.G.  
TITLE The complete nucleotide sequence of chromosome 3 of Plasmodium  
falciiparum  
JOURNALT Nature 400 (6744), 532-538 (1999)  
MEDLINE 99376085  
PUBMED 10448855  
REFERENCE 2 (bases 1 to 116696)  
AUTHORS Mungall,K., Bowman,S., Churcher,C., Lawson,D., Quail,M.,  
Rajandream,M.-A. and Barrell,B.  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 116696)  
AUTHORS Lawson,D., Bowman,S. and Barrell,B.  
TITLE Direct Submission





\* 21986 22671: contig of 686 bp in length  
\* 22672 22771: gap of 100 bp  
\* 22772 23451: contig of 680 bp in length  
\* 23452 23551: gap of 100 bp  
\* 23552 24267: contig of 716 bp in length  
\* 24268 24367: gap of 100 bp  
\* 24368 25073: contig of 706 bp in length  
\* 25074 25173: gap of 100 bp  
\* 25174 25867: contig of 694 bp in length  
\* 25868 25967: gap of 100 bp  
\* 25968 26669: contig of 702 bp in length  
\* 26670 26769: gap of 100 bp  
\* 26770 27445: contig of 676 bp in length  
\* 27446 27545: gap of 100 bp  
\* 27546 28224: contig of 679 bp in length  
\* 28225 28324: gap of 100 bp  
\* 28325 29002: contig of 678 bp in length  
\* 29003 29102: gap of 100 bp  
\* 29103 29785: contig of 683 bp in length  
\* 29786 29885: gap of 100 bp  
\* 29886 30571: contig of 686 bp in length  
\* 30572 30671: gap of 100 bp  
\* 30672 31357: contig of 686 bp in length  
\* 31358 31457: gap of 100 bp  
\* 31458 32168: contig of 711 bp in length  
\* 32169 32268: gap of 100 bp  
\* 32269 32951: contig of 683 bp in length  
\* 32952 33051: gap of 100 bp  
\* 33052 33756: contig of 705 bp in length  
\* 33757 33856: gap of 100 bp  
\* 33857 34566: contig of 710 bp in length  
\* 34567 34666: gap of 100 bp  
\* 34667 35376: contig of 710 bp in length  
\* 35377 35476: gap of 100 bp  
\* 35477 36143: contig of 667 bp in length  
\* 36144 36243: gap of 100 bp  
\* 36244 36924: contig of 681 bp in length  
\* 36925 37024: gap of 100 bp  
\* 37025 37701: contig of 677 bp in length  
\* 37702 37801: gap of 100 bp  
\* 37802 38493: contig of 692 bp in length  
\* 38494 38593: gap of 100 bp  
\* 38594 39274: contig of 681 bp in length  
\* 39275 39374: gap of 100 bp  
\* 39375 40071: contig of 697 bp in length  
\* 40072 40171: gap of 100 bp  
\* 40172 40870: contig of 699 bp in length  
\* 40871 40970: gap of 100 bp  
\* 40971 41659: contig of 689 bp in length  
\* 41660 41759: gap of 100 bp  
\* 41760 42434: contig of 675 bp in length  
\* 42435 42534: gap of 100 bp  
\* 42535 43207: contig of 673 bp in length  
\* 43208 43307: gap of 100 bp  
\* 43308 43980: contig of 673 bp in length  
\* 43981 44080: gap of 100 bp  
\* 44081 44768: contig of 688 bp in length  
\* 44769 44868: gap of 100 bp  
\* 44869 45556: contig of 688 bp in length  
\* 45557 45656: gap of 100 bp  
\* 45657 46348: contig of 692 bp in length  
\* 46349 46448: gap of 100 bp  
\* 46449 47143: contig of 695 bp in length  
\* 47144 47243: gap of 100 bp  
\* 47244 47943: contig of 700 bp in length  
\* 48043 48043: gap of 100 bp  
\* 48044 48727: contig of 684 bp in length  
\* 48728 48827: gap of 100 bp  
\* 48828 49540: contig of 713 bp in length  
\* 49541 49640: gap of 100 bp  
\* 49641 50332: contig of 692 bp in length  
\* 50333 50432: gap of 100 bp  
\* 50433 51127: contig of 695 bp in length

\* 51128 51227: gap of 100 bp  
\* 51228 51915: contig of 688 bp in length  
\* 51916 52015: gap of 100 bp  
\* 52016 52678: contig of 663 bp in length  
\* 52679 52778: gap of 100 bp  
\* 52779 53460: contig of 682 bp in length  
\* 53461 53560: gap of 100 bp  
\* 53561 54251: contig of 691 bp in length  
\* 54252 54351: gap of 100 bp  
\* 54352 54921: contig of 570 bp in length  
\* 54922 55021: gap of 100 bp  
\* 55022 55702: contig of 681 bp in length  
\* 55703 55802: gap of 100 bp

Query Match 2.9%; Score 35; DB 2: Length 149710;

Best Local Similarity 100.0%; Pred. No. 3.3e-08; Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1183 TCGTTTAAAAA 1217

Db 80812 TCGTTTAAAAA 80778

RESULT 7

AC102579

LOCUS

DEFINITION

AC102579

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC102579 Mus musculus clone RP23-248E1, WORKING DRAFT SEQUENCE, 29 unordered pieces.  
AC102579 AC102579, 2 GI:22381563  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
Mus musculus.  
house mouse.

REFERENCE 1 (bases 1 to 203610)  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
Birren,B., Nusbaum,C. and Lander,E.  
Mus musculus, clone RP23-248E1  
Unpublished  
2 (bases 1 to 203610)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Bouckgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collamore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Lakocque,K., Lamazares,R., Landers,T., Lehocsky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., MCPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Triglio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 203610)  
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Bouckgalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collamore,A., Cooke,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,

Karatas, A., Kelle, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., MacDonald, P., Major, J., Matthews, C., McCarthy, M., Melartin, J., Menus, L., Mhova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Nordu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Piere, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Teste, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A., and Zody, M.

Direct Submission  
Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Aug 21, 2002 this sequence version replaced gi:17061655.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: W1BR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

Project Information  
Center project name: L19034  
Center clone name: 248\_E.1

Summary Statistics  
Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 194595 bases at least Q40  
Consensus quality: 198128 bases at least Q30  
Consensus quality: 199686 bases at least Q20  
Insert size: 204000; agarose-fp  
Insert size: 200810; sum-of-contigs  
Quality coverage: 6.3 in Q20 bases; agarose-fp  
Quality coverage: 6.4 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 29 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1366: contig of 1366 bp in length  
1367 1466: gap of 100 bp  
1467 2393: contig of 926 bp in length  
2393 2492: gap of 100 bp  
2493 3944: contig of 1452 bp in length  
3945 4044: gap of 100 bp  
4045 5485: contig of 1441 bp in length  
5486 5585: gap of 100 bp  
5586 6815: contig of 1230 bp in length  
6816 6915: gap of 100 bp  
6916 8703: contig of 1788 bp in length  
8704 8803: gap of 100 bp  
8804 10621: contig of 1818 bp in length  
10622 10721: gap of 100 bp  
10722 13000: contig of 2279 bp in length  
13001 13100: gap of 100 bp  
13101 15018: contig of 1918 bp in length  
15019 15118: gap of 100 bp  
15119 17273: contig of 2155 bp in length  
17274 17373: gap of 100 bp  
17374 21030: contig of 3657 bp in length  
21031 21130: gap of 100 bp  
21131 24393: contig of 3263 bp in length  
24394 24493: gap of 100 bp  
24494 27284: contig of 2791 bp in length  
27285 27384: gap of 100 bp  
27385 53071: contig of 25687 bp in length

FEATURES	source
*	*
53072 53171: gap of 100 bp	
53172 56584: contig of 3413 bp in length	
56585 56684: gap of 100 bp	
56685 63948: contig of 7264 bp in length	
63949 64048: gap of 100 bp	
64049 71335: contig of 7287 bp in length	
71336 71435: gap of 100 bp	
71436 78031: contig of 6596 bp in length	
78032 78131: gap of 100 bp	
78132 86041: contig of 7910 bp in length	
86042 86141: gap of 100 bp	
86142 97779: contig of 11638 bp in length	
97780 97879: gap of 100 bp	
97880 106798: contig of 8919 bp in length	
106799 106898: gap of 100 bp	
106899 120699: contig of 14071 bp in length	
120970 121069: gap of 100 bp	
121070 129692: contig of 8623 bp in length	
129693 129792: gap of 100 bp	
129793 141332: contig of 11540 bp in length	
141333 141432: gap of 100 bp	
141433 151350: contig of 9918 bp in length	
151351 151450: gap of 100 bp	
151451 163581: contig of 12131 bp in length	
163582 163681: gap of 100 bp	
163682 178518: contig of 14837 bp in length	
178519 178618: gap of 100 bp	
178619 201081: contig of 22463 bp in length	
201082 201181: gap of 100 bp	
201182 203610: contig of 2429 bp in length.	
Location/Qualifiers	
1..203610	
/organism="Mus musculus"	
/db_xref="taxon:10090"	
/clone="RP23-248B1"	
/clone_lib="RPCI-23 Female Mouse BAC"	
1..1366	
/note="assembly_fragment"	
clone end:sp6	
vector side:left"	
1467..2392	
/note="assembly_fragment"	
2493..3944	
/note="assembly_fragment"	
4045..5485	
/note="assembly_fragment"	
5586..6815	
/note="assembly_fragment"	
6916..8703	
/note="assembly_fragment"	
8804..10621	
/note="assembly_fragment"	
10722..13000	
/note="assembly_fragment"	
13101..15018	
/note="assembly_fragment"	
15119..17273	
/note="assembly_fragment"	
17374..21030	
/note="assembly_fragment"	
21131..24393	
/note="assembly_fragment"	
24494..27284	
/note="assembly_fragment"	
27385..53071	
/note="assembly_fragment"	
53172..56584	
/note="assembly_fragment"	
56685..63948	
/note="assembly_fragment"	
64049..71335	
/note="assembly_fragment"	
71436..78031	
/note="assembly_fragment"	
78132..86041	
/note="assembly_fragment"	
86142..97779	
/note="assembly_fragment"	
97780..106798	
/note="assembly_fragment"	
106799..106898	
/note="assembly_fragment"	
106899..120699	
/note="assembly_fragment"	
120970..121069	
/note="assembly_fragment"	
121070..129692	
/note="assembly_fragment"	
129693..141332	
/note="assembly_fragment"	
141333..141432	
/note="assembly_fragment"	
141433..151350	
/note="assembly_fragment"	
151351..151450	
/note="assembly_fragment"	
151451..163581	
/note="assembly_fragment"	
163582..163681	
/note="assembly_fragment"	
163682..178518	
/note="assembly_fragment"	
178519..178618	
/note="assembly_fragment"	
178619..201081	
/note="assembly_fragment"	
201082..201181	
/note="assembly_fragment"	
201182..203610	
/note="assembly_fragment"	
203610..2429	
/note="assembly_fragment"	



misc\_feature /note="assembly\_fragment"  
78132..86041  
misc\_feature /note="assembly\_fragment"  
86142..97779  
misc\_feature /note="assembly\_fragment"  
97880..106798

Query Match 2.9%; Score 35; DB 2; Length 203610;  
Best Local Similarity 100.0%; Pred. No. 3.4e-08;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1183 TCGTTTAAAAA 57 bp DNA linear PAT 06-FEB-2001  
|||||  
Db 57211 TCGTTTAAAAA 57245

RESULT 8  
AX074093  
LOCUS AX074093 57 bp DNA linear PAT 06-FEB-2001  
DEFINITION Sequence 6 from Patent WO0104361.  
ACCESSION AX074093  
VERSION AX074093.1 GI:12710305  
KEYWORDS  
SOURCE  
ORGANISM  
synthetic construct.  
artificial sequences.  
REFERENCE 1 (bases 1 to 57)  
AUTHORS Bee,G.G., Yang,Y.Y., Kolk,D.P., Giachetti,C. and McDonough,S.H.  
TITLE Detection of hiv-1 by nucleic acid amplification  
JOURNAL Patent: WO 0104361-A 6 18-JAN-2001;  
Gen-Probe Incorporated (US); Bee, Gary G. (US); Yang, Yeasing Y.  
(US); Kolk, Dan P. (US); Giachetti, Cristina (US); McDonough,  
Sherrol Hoffa (US)

FEATURES  
source  
1..57  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="synthetic capture oligomer with 3' tail sequence"  
BASE COUNT 35 a 8 c 4 g 10 t  
ORIGIN

Query Match 2.8%; Score 34; DB 6; Length 57;  
Best Local Similarity 100.0%; Pred. No. 9e-08;  
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1184 CGTTTAAAAA 1217  
|||||  
Db 23 CGTTTAAAAA 56

RESULT 9  
AX078311  
LOCUS AX078311 57 bp DNA linear PAT 22-FEB-2001  
DEFINITION Sequence 7 from Patent WO0107661.  
ACCESSION AX078311  
VERSION AX078311.1 GI:13158002  
KEYWORDS  
SOURCE  
ORGANISM  
synthetic construct.  
artificial sequences.  
REFERENCE 1 (bases 1 to 57)  
AUTHORS Nunomura,K.  
TITLE Polynucleotide amplification method  
JOURNAL Patent: WO 0107661-A 7 01-FEB-2001;  
Gen-Probe Incorporated (US)

FEATURES  
source  
1..57  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="Sequence of the (-)4258 A30 capture probe."  
BASE COUNT 35 a 8 c 4 g 10 t  
ORIGIN

Query Match 2.8%; Score 34; DB 6; Length 57;  
Best Local Similarity 100.0%; Pred. No. 9e-08;  
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1184 CGTTTAAAAA 1217  
|||||  
Db 23 CGTTTAAAAA 56

RESULT 10  
AX397797  
LOCUS AX397797 57 bp DNA linear PAT 27-MAY-2002  
DEFINITION Sequence 31 from Patent WO0220852.  
ACCESSION AX397797  
VERSION AX397797.1 GI:21260671  
KEYWORDS  
SOURCE  
ORGANISM  
synthetic construct.  
artificial sequences.

REFERENCE 1  
AUTHORS Yang,Y.Y., Brentano,S.T., Babola,O., Tran,N. and Vernet,G.  
TITLE Amplification of hiv-1 sequences for detection of sequences  
associated with drug-resistance mutations  
JOURNAL Patent: WO 0220852-A 31 14-MAR-2002;  
Gen-Probe Incorporated Patent Dept (US); Biomerieux S.A. (FR)  
FEATURES  
source  
1..57  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="Capture oligonucleotide"  
BASE COUNT 35 a 8 c 4 g 10 t  
ORIGIN

Query Match 2.8%; Score 34; DB 6; Length 57;  
Best Local Similarity 100.0%; Pred. No. 9e-08;  
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1184 CGTTTAAAAA 1217  
|||||  
Db 23 CGTTTAAAAA 56

RESULT 11  
AV119210  
LOCUS AV119210 540 bp mRNA linear INV 16-JUN-2002  
DEFINITION Drosophila melanogaster S015809 full insert cDNA.  
ACCESSION AV119210  
VERSION AV119210.1 GI:21430783  
KEYWORDS  
SOURCE  
ORGANISM  
fruit fly.  
Drosophila melanogaster

REFERENCE  
AUTHORS  
Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,  
Champe,M., Chavez,C., Dorsett,V., Dresnek,D., Farfan,D., Frise,E.,  
George,R., Gonzalez,M., Guarin,H., Kronmiller,B., Li,P., Liao,G.,  
Miranda,A., Mungall,C.J., Nunco,J., Pacleb,J., Paragas,V., Park,S.,  
Patel,S., Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Rubin,G.M.  
and Celniker,S.  
Direct Submission  
Submitted (07-JUN-2002) Berkeley Drosophila Genome Project,  
Lawrence Berkeley National Laboratory, One Cyclotron Road,  
Berkeley, CA 94720, USA  
Sequence submitted by:  
Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory  
Berkeley, CA 94720

TITLE  
JOURNAL  
COMMENT  
This clone was sequenced as part of a high-throughput process to  
sequence clones from Drosophila Gene Collection 1 (Rubin et al.,  
Science 2000). The sequence has been subjected to integrity checks  
for sequence accuracy, presence of a polyA tail and contiguity

within 100 kb in the genome. Thus we believe the sequence to reflect accurately this particular cDNA clone. However, there are artifacts associated with the generation of cDNA clones that may have not been detected in our initial analyses such as internal priming, priming from contaminating genomic DNA, retained introns due to reverse transcription of unspliced precursor RNAs, and reverse transcriptase errors that result in single base changes. For further information about this sequence, including its location and relationship to other sequences, please visit our web site (<http://fruitfly.berkeley.edu>) or send email to [cdna@fruitfly.berkeley.edu](mailto:cdna@fruitfly.berkeley.edu).

## FEATURES

## SOURCE

1. 540  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
1. 539  
misc\_feature  
/note="sim4 alignment with AE003815 (band-50D7-band-50P6)"  
52. 441  
CDS  
/note="Longest ORF"  
/codon\_start=1  
/product="SD15809P"  
/protein\_id="AAM51070.1"  
/db\_xref="GI:21430784.1"  
/translation="MSLSPPASGIGRFYAKSAKIRFVRLCCTRPPIHYIVERRKNO  
HQPVEGVGSDPLNDYNERLVANLTERIRYWLKGANLSTPAELLGIALPIHP  
RTYMTANRNRTAAEAASPEKAEKSTA"

BASE COUNT 177 a 132 c 118 g 113 t  
ORIGIN

Query Match 2.8%; Score 34; DB 3; Length 540;  
Best Local Similarity 100.0%; Pred. No. 9.8e-08;

Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1184 CGTTTAAAAAAAAAAAAAAAAAAAAA 1217  
|||||  
Db 482 CGTTTAAAAAAAAAAAAAAAAAAAAA 515

## RESULT 12

## LOCUS

AY119449 853 bp mRNA linear INV 15-JUN-2002  
DEFINITION Drosophila melanogaster AT11913 full insert cDNA.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

AY119449 853 bp mRNA linear INV 15-JUN-2002  
DEFINITION Drosophila melanogaster AT11913 full insert cDNA.  
ACCESSION AY119449  
VERSION AY119449.1 GI:21428767  
KEYWORDS F11 CDNA.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 853)  
Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Carlson, J.,  
Champe, M., Chavez, C., Dorsett, V., Dresnek, D., Farfan, D., Frise, E.,  
Gonzalez, M., Guarini, H., Krommiller, B., Li, P., Liao, G.,  
Miranda, A., Mungall, C. J., Nuno, J., Pacleb, J., Paragas, V., Park, S.,  
Patel, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S. E., Rubin, G. M.,  
and Ceiniker, S.  
Direct Submission  
Submitted (10-JUN-2002) Berkeley Drosophila Genome Project,  
Lawrence Berkeley National Laboratory, One Cyclotron Road,  
Berkeley, CA 94720, USA  
Sequence submitted by:  
Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory  
Berkeley, CA 94720  
This clone was sequenced as part of a high-throughput process to  
sequence clones from Drosophila Gene Collection 1 (Rubin et al.,  
Science 2000). The sequence has been subjected to integrity checks  
for sequence accuracy, presence of a polyA tail and contiguity  
within 100 kb in the genome. Thus we believe the sequence to  
reflect accurately this particular cDNA clone. However, there are  
artifacts associated with the generation of cDNA clones that may  
have not been detected in our initial analyses such as internal

priming, priming from contaminating genomic DNA, retained introns  
due to reverse transcription of unspliced precursor RNAs, and  
reverse transcriptase errors that result in single base changes.  
For further information about this sequence, including its location  
and relationship to other sequences, please visit our web site  
(<http://fruitfly.berkeley.edu>) or send email to  
[cdna@fruitfly.berkeley.edu](mailto:cdna@fruitfly.berkeley.edu).

## FEATURES

## SOURCE

1. 853  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
1. 760  
misc\_feature  
/note="sim4 alignment with AE003598 (band-79E4-band-79F3)"  
71. 508  
CDS  
/note="Longest ORF"  
/codon\_start=1  
/product="AT11913P"  
/protein\_id="AAM50103.1"  
/db\_xref="GI:21428768"  
/translation="MPKPKHNDVAGFLSRVNFILGRTHTAHRADWVSPRTQPP  
NIPSGPTSLFANYTYRDPRLVYKFDVYVQEHKMLTKAYKEEAKAKQAKSGDA  
PXDSPFIPVGPXTDTNECDENSGAKLPTPKVSWSEPH"

BASE COUNT 302 a 208 c 179 g 164 t  
ORIGIN

Query Match 2.8%; Score 34; DB 3; Length 853;  
Best Local Similarity 100.0%; Pred. No. 1e-07;

Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1184 CGTTTAAAAAAAAAAAAAAAAAAAAA 1217  
|||||  
Db 754 CGTTTAAAAAAAAAAAAAAAAAAAAA 787

## RESULT 13

## LOCUS

AY060916 1551 bp mRNA linear INV 08-NOV-2001  
DEFINITION Drosophila melanogaster GM06077 full length cDNA.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

AY060916 1551 bp mRNA linear INV 08-NOV-2001  
DEFINITION Drosophila melanogaster GM06077 full length cDNA.  
ACCESSION AY060916  
VERSION AY060916.1 GI:16768489  
KEYWORDS F11 CDNA.  
SOURCE Drosophila melanogaster.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 1551)  
Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Carlson, J.,  
Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George, R.,  
Gonzalez, M., Guarini, H., Li, P., Liao, G., Miranda, A., Mungall, C. J.,  
Nuno, J., Pacleb, J., Paragas, V., Park, S., Phouanavong, S., Wan, K.,  
Yu, C., Lewis, S. E., Rubin, G. M., and Ceiniker, S.  
Direct Submission  
Submitted (30-OCT-2001) Berkeley Drosophila Genome Project,  
Lawrence Berkeley National Laboratory, One Cyclotron Road,  
Berkeley, CA 94720, USA  
Sequence submitted by:  
Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory  
Berkeley, CA 94720  
This clone was sequenced as part of a high-throughput process to  
sequence clones from Drosophila Gene Collection 1 (Rubin et al.,  
Science 2000). The sequence has been subjected to integrity checks  
for sequence accuracy, presence of a polyA tail and contiguity  
within 100 kb in the genome. Thus we believe the sequence to  
reflect accurately this particular cDNA clone. However, there are  
artifacts associated with the generation of cDNA clones that may  
have not been detected in our initial analyses such as internal  
priming, priming from contaminating genomic DNA, retained introns  
due to reverse transcription of unspliced precursor RNAs, and  
reverse transcriptase errors that result in single base changes.  
For further information about this sequence, including its location  
and relationship to other sequences, please visit our web site

(http://fruitfly.berkeley.edu) or send email to  
cdna@fruitfly.berkeley.edu.

FEATURES  
source

1. 1551  
Location/Qualifiers  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/map="97A4-97A4"  
1. 1551  
/gene="CG6073"  
/note="alignment with genomic scaffold AE003755"  
/db\_xref="FLYBASE:FBgn0039417"  
59. 1168  
/gene="CG6073"  
/note="Longest ORF"  
/codon\_start=1  
/product="GM06077p"  
/protein\_id="AAL28464.1"  
/db\_xref="GI:16768490"  
/db\_xref="FLYBASE:FBgn0039417"  
/translation="METVKELVQPNQNRDLKVALTHVLGTLGSSEKSAIISLD  
EMLAIKGLFDANOTVAKDAVLSILNLTSEEAALKVQLAKQLQPPPAIVEAAKE  
ITNQSLADPWSMVLNSLTRVESLVHILDTLERRDHLPLAKAPALQDYNKKAK  
LHYAPIFCMLTQVSRGRELCHRRYELLEKLFPFASFGSVVRGGTIGLKNVCFD  
TVYHDVILNQSSILVAIQLCPPEPESDEDNLPIELQYLPESKRTREDDPLRM  
LLECLQLCSTRRSRILRSRGVIELFRKWEAKVGDCKDCLLACENVVDIILIKKE  
BEIGLDNYKTEVEVPAEQSEKFQVEDAAVYKSLID"

gene

CDS

59. 1168

/db\_xref="FLYBASE:FBgn0039417"

/note="alignment with genomic scaffold AE003755"

CDS

98. 1417

/note="Vector: pBluescript"

/lab\_host="DH10B"  
/clone\_lib="NIH MGC 97"  
/tissue\_type="Testis"  
/clone="MGC:41867 IMAGE:5265994"  
/db\_xref="taxon:9606"  
/organism="Homo sapiens"  
1. 1627  
Location/Qualifiers

This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 4557454.

FEATURES  
source

1. 1627  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="MGC:41867 IMAGE:5265994"  
/tissue\_type="Testis"  
/clone\_lib="NIH MGC 97"  
/lab\_host="DH10B"  
/note="Vector: pBluescript"  
98. 1417  
/codon\_start=1  
/product="Similar to choline kinase"  
/protein\_id="AAH36471.1"  
/db\_xref="GI:22209091"  
/translation="MKTKFTCTGGEAPSPGLGLLSCGSSAAPAGVGQORDAASDLE  
SKQLGGQPPALPPLPPPLPLPLPQPPPPQPPADPEQPERTRRRAYLWCKEFLFGA  
WGLRDEDFHISVIRGSLNMFQCSLPTDTATLGDPEPKVLLRLYGAILQVGAEMV  
LRSVMFATLBSRGLPKYIGIPQGRLEQIPSRRLDTEELGLPDISAEIAEKWATFH  
GNMPPNKEPKWLGMEKYLEVLRIKETSESRKKLKLKLSYNLPLELENLSLLE  
STPSVFECHNDCEGNILLEGRENSEKQKMLIDFEYSYNYRGDFDIGNHCEHMY  
DYSIKYFFFRANIKPTTKKQLHPFISILPAFONDENLSEKSSIKKEEMLLLEV  
RPAASHFLWQMSIVQAKLSISIFGMDYQAQAFDAYFHOKRKLGV"

BASE COUNT  
ORIGIN

422 a 419 c 432 g 354 t

Query Match

Best Local Similarity 2.8%; Score 34; DB 9; Length 1627;

Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1184

CGTTTAAAAA 1217

Db 1580

CGTTTAAAAA 1613

RESULT 15  
AX070713

LOCUS

AX070713 Drosophila melanogaster SD07518 full length cDNA. 1771 bp mRNA linear INV 20-DEC-2001

DEFINITION

AX070713 Drosophila melanogaster SD07518 full length cDNA.

ACCESSION

AX070713 AY070713.1 GI:17944593

VERSION

AX070713.1 GI:17944593

KEYWORDS

FLI CDNA.

SOURCE

Drosophila melanogaster.

ORGANISM

Drosophila melanogaster.

REFERENCE

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

AUTHORS

Ephyrdoidea; Drosophilidae; Drosophila.

1 (bases 1 to 1771)

Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Carlson, J.,

Champe, M., Chavez, C., Dorsett, V., Dresnek, D., Farfan, D., Frise, E.,

George, R., Gonzalez, M., Guarin, H., Kronmiller, B., Li, P., Liao, G.,

Miranda, A., Mungall, C.J., Nunco, J., Pacieb, J., Paragas, V., Park, S.,

Patel, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Rubin, G.M.

and Celniker, S.

Direct Submission

Submitted (18-DEC-2001) Berkeley Drosophila Genome Project,

Lawrence Berkeley National Laboratory, One Cyclotron Road,

Berkeley, CA 94720, USA

Sequence submitted by:

Berkeley Drosophila Genome Project

Lawrence Berkeley National Laboratory

Berkeley, CA 94720

COMMENT

This clone was sequenced as part of a high-throughput process to  
sequence clones from Drosophila Gene Collection 1 (Rubin et al.,  
Science 2000). The sequence has been subjected to integrity checks  
for sequence accuracy, presence of a polyA tail and contiguity  
within 100 kb in the genome. Thus we believe the sequence to  
reflect accurately this particular cDNA clone. However, there are  
artifacts associated with the generation of cDNA clones that may  
have not been detected in our initial analyses such as internal  
priming, priming from contaminating genomic DNA, retained introns  
due to reverse transcription of unspliced precursor RNAs, and

FEATURES  
source

1. 1551  
Location/Qualifiers  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/map="97A4-97A4"  
1. 1551  
/gene="CG6073"  
/note="alignment with genomic scaffold AE003755"  
/db\_xref="FLYBASE:FBgn0039417"  
59. 1168  
/gene="CG6073"  
/note="Longest ORF"  
/codon\_start=1  
/product="GM06077p"  
/protein\_id="AAL28464.1"  
/db\_xref="GI:16768490"  
/db\_xref="FLYBASE:FBgn0039417"  
/translation="METVKELVQPNQNRDLKVALTHVLGTLGSSEKSAIISLD  
EMLAIKGLFDANOTVAKDAVLSILNLTSEEAALKVQLAKQLQPPPAIVEAAKE  
ITNQSLADPWSMVLNSLTRVESLVHILDTLERRDHLPLAKAPALQDYNKKAK  
LHYAPIFCMLTQVSRGRELCHRRYELLEKLFPFASFGSVVRGGTIGLKNVCFD  
TVYHDVILNQSSILVAIQLCPPEPESDEDNLPIELQYLPESKRTREDDPLRM  
LLECLQLCSTRRSRILRSRGVIELFRKWEAKVGDCKDCLLACENVVDIILIKKE  
BEIGLDNYKTEVEVPAEQSEKFQVEDAAVYKSLID"

gene

CDS

59. 1168

/db\_xref="FLYBASE:FBgn0039417"

/note="alignment with genomic scaffold AE003755"

CDS

98. 1417

/note="Vector: pBluescript"

/lab\_host="DH10B"  
/clone\_lib="NIH MGC 97"  
/tissue\_type="Testis"  
/clone="MGC:41867 IMAGE:5265994"  
/db\_xref="taxon:9606"  
/organism="Homo sapiens"  
1. 1627  
Location/Qualifiers

This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 4557454.

FEATURES  
source

1. 1627  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="MGC:41867 IMAGE:5265994"  
/tissue\_type="Testis"  
/clone\_lib="NIH MGC 97"  
/lab\_host="DH10B"  
/note="Vector: pBluescript"  
98. 1417  
/codon\_start=1  
/product="Similar to choline kinase"  
/protein\_id="AAH36471.1"  
/db\_xref="GI:22209091"  
/translation="MKTKFTCTGGEAPSPGLGLLSCGSSAAPAGVGQORDAASDLE  
SKQLGGQPPALPPLPPPLPLPQPPPPQPPADPEQPERTRRRAYLWCKEFLFGA  
WGLRDEDFHISVIRGSLNMFQCSLPTDTATLGDPEPKVLLRLYGAILQVGAEMV  
LRSVMFATLBSRGLPKYIGIPQGRLEQIPSRRLDTEELGLPDISAEIAEKWATFH  
GNMPPNKEPKWLGMEKYLEVLRIKETSESRKKLKLKLSYNLPLELENLSLLE  
STPSVFECHNDCEGNILLEGRENSEKQKMLIDFEYSYNYRGDFDIGNHCEHMY  
DYSIKYFFFRANIKPTTKKQLHPFISILPAFONDENLSEKSSIKKEEMLLLEV  
RPAASHFLWQMSIVQAKLSISIFGMDYQAQAFDAYFHOKRKLGV"

BASE COUNT  
ORIGIN

422 a 419 c 432 g 354 t

Query Match

Best Local Similarity 2.8%; Score 34; DB 9; Length 1627;

Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1184

CGTTTAAAAA 1217

Db 1518

CGTTTAAAAA 1551

RESULT 14  
BC036471

LOCUS

BC036471 Homo sapiens, Similar to choline kinase, clone MGC:41867 1627 bp mRNA linear PRI 12-AUG-2002

DEFINITION

BC036471 Homo sapiens, Similar to choline kinase, clone MGC:41867

ACCESSION

BC036471 IMAGE:5265994, mRNA, complete cds.

VERSION

BC036471.1 GI:22209090

KEYWORDS

MGC.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1627)

Strausberg, R.

Direct Submission

Submitted (09-AUG-2002) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: http://mgc.nci.nih.gov

Contact: MGC help desk

Email: cgapbs@mail.nih.gov

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki

Toshiyuki and Piero Carninci (RIKEN)

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Institute for Systems Biology

http://www.systemsbio.org

Contact: amadan@systemsbiology.org

Anup Madan, Jessica Fahey, Brin Helton, Mark Ketterman, Anuradha

Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov

Series: IRAK Plate: 73 Row: n Column: 14

reverse transcriptase errors that result in single base changes.  
For further information about this sequence, including its location  
and relationship to other sequences, please visit our web site  
(<http://fruitfly.berkeley.edu>) or send email to  
[cdna@fruitfly.berkeley.edu](mailto:cdna@fruitfly.berkeley.edu).

FEATURES  
source Location/Qualifiers

gene 1..1771  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/map="21C5-21C5"

CDS 1..1771  
/gene="shanti"  
/note="alignment with genomic scaffold AE003589"  
/db\_xref="FLYBASE:FBgn0031260"  
85..1254  
/gene="shanti"  
/note="Longest ORF"  
/product="SD07518p"  
/protein\_id="AAI48184.1"  
/db\_xref="GI:1794594"  
/db\_xref="FLYBASE:FBgn0031260"

BASE COUNT 465 a 464 c 434 g 408 t  
ORIGIN  
/translation="MAEEVIGTKEVLEKGIENVNAPKNEAPGEKPESTPEGMAYAY  
SSLVMMALPILFGSIRSVKHLKKSSTGEKADMTKKDAMPPLIASPALFGLYLFF  
KIPQKHINYLITGYFEVGLVLAHLSPVINSIMPAVAVKVPFHLPTKGEGKHE  
DIVVKKSTHDIVCLVSSAIGWYLLKKHIANNLFGLAFAINGVEMHLNPNVGV  
ILSGLEFYDIPWVPGTNTMVAKSPAPIKLYFPDILLENGANSNFAMIGLDIV  
IPGIFILLLRPDSKKRRTIRYFSTLIATFDGLKATITVNHVFKHQPALLIVPA  
CMGIFPLVALIRGELKLVLFAYEDHPERKEKKEKDEGVSSSGSKKESKKGK"

Query Match 2.8%; Score 34; DB 3; Length 1771;  
Best Local Similarity 100.0%; Pred. No. 1e-07;  
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1184 CGTTTAAAAAAAAAAAAAAAAAAAAA 1217  
Db 1705 CGTTTAAAAAAAAAAAAAAAAAAAAA 1738

Search completed: December 17, 2002, 18:50:05  
Job time : 3745 secs



P1 Falco SC, Allen SM, Anderson SL;  
 XX WPI; 2000-182430/16.  
 DR P-PSDB; AA44789.  
 XX  
 PT New nucleic acid molecule and chimeric gene encoding an adenosine-5'-  
 PT phosphosulphate kinase, useful for altering expression of sulfate  
 PT assimilation protein in plants  
 XX  
 PS Claim 3; Page 30; 42p; English.  
 XX  
 CC The present sequence is a cDNA encoding corn adenylylsulphate kinase  
 CC (APS kinase), also known as adenosine-5' phosphosulphate kinase. This is  
 CC obtained from clone p0016.ctscf40rb, derived from corn pooled tassel  
 CC shoots, p0016 cDNA library. APS kinase is a sulphate assimilation  
 CC protein, that catalyses the conversion of adenosine-5' phosphosulphate  
 CC (APS) to 3'-phospho-adenosine-5' phosphosulphate (PAPS). This sequence  
 CC is used as probe and primers to identify, obtain and synthesise sulphate  
 CC assimilation proteins from other plants. It is also used to produce  
 CC transgenic plants, that are useful for altering the expression levels of  
 CC a sulphate assimilation protein. The APS kinase peptides are useful for  
 CC producing antibodies, that are used to screen and isolate cDNA clones.  
 XX  
 SQ Sequence 1217 BP; 344 A; 292 C; 303 G; 278 T; 0 other;  
 Query Match 100.0%; Score 1217; DB 21; Length 1217;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GCGTCGGTTTCATTCATCATCAACAGAACTCTGTGTCAACACACGGACGACCC 60  
 Db 1 GCGTCGGTTTCATTCATCAATCAACAGAACTCTGTGTCAACACACGACCAACCC 60  
 QY 61 GAGCCAGAGCCCGGCCAGCAGCCAGGCGCAACGGCAAGGCAACCTCTCCAGCCC 120  
 Db 61 GAGCCAGAGCCCGGCCAGCAGCCAGGCGCAACGGCAAGGCAACCTCTCCAGCCC 120  
 QY 121 GAGCCGAGCGCTCGCGTCATCTCTGTAATCCACAGCGCGGCTCCCGTCTCCAGG 180  
 Db 121 GAGCCGAGCGCTCGCGTCATCTCTGTAATCCACAGCGCGGCTCCCGTCTCCAGG 180  
 QY 181 CCTCACCCCTTACGATGCGCACTCCGCGCTCGATTCATCGATCGCTCACTCCCGTTC 240  
 Db 181 CCTCACCCCTTACGATGCGCACTCCGCGCTCGATTCATCGATCGCTCACTCCCGTTC 240  
 QY 241 CTGACACTCTTCCGCGGCTCTCCGCACTGATGTGGGCGCCCGAGGGGAGGCGCCGG 300  
 Db 241 CTGACACTCTTCCGCGGCTCTCCGCACTGATGTGGGCGCCCGAGGGGAGGCGCCGG 300  
 QY 301 TGGCGTACGCACTCCACCGCGGCAATTGGGCGGTGTGCGCGCGCGCGCGGAGATGG 360  
 Db 301 TGGCGTACGCACTCCACCGCGGCAATTGGGCGGTGTGCGCGCGCGCGGAGATGG 360  
 QY 361 AGCAGCGCCCGGGAGGCGCCCGCACACCCACAGTGAAGAGAAAGCTGTAAATGTGAACAT 420  
 Db 361 AGCAGCGCCCGGGAGGCGCCCGCACACCCACAGTGAAGAGAAAGCTGTAAATGTGAACAT 420  
 QY 421 TGGGAATCGACTAATATTTATGTGGCAAAATTCCTGATGGACAATCTGATAGACAA 480  
 Db 421 TGGGAATCGACTAATATTTATGTGGCAAAATTCCTGATGGACAATCTGATAGACAA 480  
 QY 481 ATTGCTGGAGCAAAAAGGCTGTGTCTGATGATGAACAGACTCAATGCTTACGAGAAAG 540  
 Db 481 ATTGCTGGAGCAAAAAGGCTGTGTCTGATGATGAACAGACTCAATGCTTACGAGAAAG 540  
 QY 541 TACTCTTGATGTGACCTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 600  
 Db 541 TACTCTTGATGTGACCTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 600  
 QY 601 TGATGTGACAACTTCAGACATGGCTTAATAGATTTTAACCTTTAAGGACAGAACCG 660  
 Db 601 TGATGTGACAACTTCAGACATGGCTTAATAGATTTTAACCTTTAAGGACAGAACCG 660

QY 661 TGCAGAAAATATGACGAAGTGTGGAAGTGCCAAACCTTTTGTGATGCTGTGTGAT 720  
 Db 661 TGCAGAAAATATGACGAAGTGTGGAAGTGCCAAACCTTTTGTGATGCTGTGTGAT 720  
 QY 721 ATGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780  
 Db 721 ATGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780  
 QY 781 TCCACATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840  
 Db 781 TCCACATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840  
 QY 841 TGCATCAAGGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900  
 Db 841 TGCATCAAGGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900  
 QY 901 TGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960  
 Db 901 TGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960  
 QY 961 ATGCGCTTCAACCAAGCAAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1020  
 Db 961 ATGCGCTTCAACCAAGCAAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1020  
 QY 1021 TTTGCAAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
 Db 1021 TTTGCAAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
 QY 1081 TGTGACACAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140  
 Db 1081 TGTGACACAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140  
 QY 1141 TTAACAGAGTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200  
 Db 1141 TTAACAGAGTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200  
 QY 1201 AAAAAAAAAAAAAAAAAA 1217  
 Db 1201 AAAAAAAAAAAAAAAAAA 1217  
 RESULT 2  
 AAC99029  
 ID AAC99029 standard; cDNA; 2952 BP.  
 XX  
 AC AAC99029;  
 XX  
 DT 09-MAR-2001 (first entry)  
 XX  
 DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:257.  
 XX  
 KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;  
 KW detection; diagnosis; identification; cytostatic; neuroprotective;  
 KW neoplastic; immunomodulatory; relaxant; contraceptive; gynaecological;  
 KW antineoplastic; cardiant; gene therapy; chromosome mapping;  
 KW linkage analysis; tissue identification; tissue typing; forensic;  
 KW neutral; immune system; muscular; reproductive; gastrointestinal;  
 KW pulmonary; cardiovascular; renal; proliferative; 88.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200055320-A1.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PF 08-MAR-2000; 2000WO-US05989.  
 XX  
 PR 12-MAR-1999; 99US-0124270.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM;  
 XX

DR WPI; 2000-579444/54.  
XX P-PSDB; AAB54264.  
PT New nucleic acid that is a pancreatic cancer antigen for preventing,  
PT treating, or ameliorating a medical condition, particular pancreatic  
PT cancer, or for use in assays for diagnosing a pathological condition -  
XX  
XX Claim 1; Page 696-697; 1379pp; English.  
XX  
XX AAC98773 to AAC99231 encode the human pancreatic cancer associated  
CC proteins, called pancreatic cancer antigens, given in AAB54008 to  
CC AAB54466. The human pancreatic cancer antigens have cytostatic,  
CC neuroprotective, nontoxic, immunomodulatory, relaxant, contraceptive,  
CC gynaecological, cardiant and antiinflammatory activities, and can be used  
CC in gene therapy. The polynucleotide and proteins can be used for  
CC preventing, treating, or ameliorating a medical condition or in assays  
CC for diagnosing a pathological condition or a susceptibility to one in a  
CC subject. Binding partners to the proteins and the activity of the  
CC proteins can be identified. The pancreatic cancer antigens can be used to  
CC detect, treat or prevent pancreatic disorders, especially cancer.  
CC Agonists and antagonists to the antigens can be screened for. The  
CC pancreatic cancer antigen polynucleotides can be used to design nucleic  
CC acid hybridisation probes that can be used in chromosome mapping, linkage  
CC analysis, tissue identification and/or typing and a variety of forensic  
CC and diagnostic methods. The proteins can be used to generate antibodies  
CC which are used to purify, detect and target the polypeptides, including  
CC both in vivo and in vitro diagnostic and therapeutic methods. The  
CC proteins can be used to treat or prevent neural, immune system, muscular,  
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or  
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent  
CC sequences used in the exemplification of the present invention.  
XX  
XX Sequence 2952 BP; 766 A; 645 C; 671 G; 868 T; 2 other;  
SQ  
Query Match 3.0%; Score 36; DB 21; Length 2952;  
Best Local Similarity 100.0%; Pred. No. 9e-05;  
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy .1182 TTCGTTTAAAAA 1217  
Db 2916 TTCGTTTAAAAA 2951  
RESULT 3  
ABV57020  
ID ABV57020 standard; cDNA; 472 BP.  
XX  
XX AC ABV57020;  
XX  
XX DT 17-SEP-2002 (first entry)  
XX  
XX DE Human prostate expression marker cDNA 57011.  
XX  
XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
XX pharmacogenomic marker; gene; ss.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO200160860-A2.  
XX  
XX PD 23-AUG-2001.  
XX  
XX PF 20-FEB-2001; 2001WO-US05171.  
XX  
XX PR 17-FEB-2000; 2000US-183319P.  
XX  
XX PR 16-MAR-2000; 2000US-189862P.  
XX  
XX PR 25-MAY-2000; 2000US-207454P.  
XX  
XX PR 09-JUN-2000; 2000US-211314P.  
XX  
XX PR 18-JUL-2000; 2000US-219007P.  
XX  
XX PR 13-DEC-2000; 2000US-255281P.  
XX  
XX FA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX

PI Schlegel R, Endege WO, Monahan JE;  
XX  
XX WPI; 2001-662795/76.  
XX  
XX Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer -  
XX  
XX Claim 1; Page 10979; 11750pp; English.  
XX  
XX The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX  
XX Sequence 472 BP; 172 A; 103 C; 96 G; 101 T; 0 other;  
SQ  
Query Match 2.9%; Score 35; DB 23; Length 472;  
Best Local Similarity 100.0%; Pred. No. 0.00031;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1183 TCGTTTAAAAA 1217  
Db 45 TCGTTTAAAAA 79  
RESULT 4  
AAF60546  
ID AAF60546 standard; DNA; 57 BP.  
XX  
XX AC AAF60546;  
XX  
XX DT 27-APR-2001 (first entry)  
XX  
XX DE Probe #3 used in a method for quantifying analyte polynucleotides.  
XX  
XX KW Probe; HIV; ss.  
XX  
XX OS Human immunodeficiency virus type 1.  
XX  
XX PN WO200107661-A2.  
XX  
XX PD 01-FEB-2001.  
XX  
XX PF 21-JUL-2000; 2000WO-US20034.  
XX  
XX PR 23-JUL-1999; 99US-0145432.  
XX  
XX PA (GENP-) GEN-PROBE INC.  
XX  
XX PI Nunomura K;  
XX  
XX WPI; 2001-182804/18.  
XX  
XX Detecting and quantitating analyte polynucleotide in a sample, by  
PT co-amplifying analyte polynucleotide with predetermined amount of  
PT pseudo target, producing amplification products and quantifying analyte  
PT amplicons -  
XX  
XX Example 3; Page 39; 78pp; English.  
XX  
XX The present invention relates to a method for quantifying analyte  
CC

CC polynucleotides (AP). The method comprises combining a test sample of AP  
 CC with predetermined amount of a pseudo target (PT) and co-amplifying, to  
 CC produce a collection of amplification products including an analyte  
 CC amplicon if the sample contained AP and PT amplicon. The analyte amplicon  
 CC is quantified without reference to the amount of PT amplicon and related  
 CC to AP in the sample. The present sequence is a probe used in the method  
 CC of the present invention.

XX Sequence 57 BP; 35 A; 8 C; 4 G; 10 T; 0 other;

Query Match 2.8%; Score 34; DB 22; Length 57;

Best Local Similarity 100.0%; Pred. No. 0.0011; Mismatches 0; Indels 0; Gaps 0;

Oy 1184 CGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1217  
 Db 23 CGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 56

RESULT 5  
 AAF56538 ID AAF56538 standard; DNA; 57 BP.

XX AAF56538;

DT 18-APR-2001 (first entry)

DE HIV-1 detection PCR primer SEQ ID NO: 6.

KM HIV-1 detection; diagnosis; blood screening; PCR primer; probe; ss.

OS Human immunodeficiency virus type 1.

PN WQ200104361-A2.

PD 18-JAN-2001.

PF 07-JUL-2000; 2000MO-US18695.

PR 09-JUL-1999; 99US-0143072.

PA (GENP-) GEN-PROBE INC.

PA (BERG/) BEF G G.

PA (YANG/) YANG Y Y.

PA (KOLK/) KOLK D P.

PA (GIAC/) GIACHETTI C.

PA (MCDO/) MCDONOUGH S H.

PI Bee GG, Yang YY, Kolk DP, Giachetti C, McDonough SH;

DR WPI; 2001-147200/15.

PT Detecting HIV-1 nucleic acids in biological samples useful for

CC diagnosing HIV-1 infection involves using nucleic acid capture

CC oligomers, amplification oligomers and probe oligomers

PS Claim 1; Page 50; 60pp; English.

CC The present invention provides probes and PCR primers for use in the

CC detection of HIV-1. These are shown in AAF5533-AAF5589. They can be

CC used to diagnose HIV infection and to ensure that blood and blood

CC products do not contain the virus, thus enabling the prevention of HIV

CC infection during blood transfusions.

XX Sequence 57 BP; 35 A; 8 C; 4 G; 10 T; 0 other;

Query Match 2.8%; Score 34; DB 22; Length 57;

Best Local Similarity 100.0%; Pred. No. 0.0011; Mismatches 0; Indels 0; Gaps 0;

Oy 1184 CGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1217  
 Db 23 CGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 56

RESULT 6  
 AAK53131 ID AAK53131 standard; DNA; 57 BP.

XX AAK53131;

DT 12-AUG-2002 (first entry)

DE HIV detection method capture oligonucleotide #2.

KM HIV, human immunodeficiency virus; ss; probe; gag; pol;

KM protease; reverse transcriptase; infection.

OS Unidentified.

PN US2002055095-A1.

PD 09-MAY-2002.

PF 31-AUG-2001; 2001US-0944036.

PR 01-SEP-2000; 2000US-229790P.

PA (YANG/) YANG Y Y.

PA (BREN/) BRENTANO S T.

PA (BABO/) BABOLA O.

PA (TRAN/) TRAN N.

PA (VERN/) VERNET G.

PI Yang YY, Brentano ST, Babola O, Tran N, Vernet G;

DR WPI; 2002-462902/49.

PT New nucleic acid oligomers for amplifying a nucleotide sequence from

PT HIV-1 and probes for detecting the amplified product are specific for

PT gag and pol regions and are useful to detect different subtypes of

PT HIV-1

PS Example 4; Page 23; 37pp; English.

CC This invention relates to a series of nucleic acid oligomers for

CC amplifying and detecting a nucleotide sequence of human immunodeficiency

CC virus type 1 (HIV-1). The invention also comprises a labeled

CC oligonucleotide that specifically hybridizes to an HIV-1 sequence

CC derived from gag or pol sequences, having one of the sequences fully

CC defined in the specification, and a method for detecting HIV-1 in a

CC biological sample, comprising mixing the sample with two or more of the

CC amplification oligomers that specifically amplify at least one HIV-1

CC target sequence within gag and a pol sequence which is a protease or

CC reverse transcriptase sequence, amplifying the target, and detecting the

CC amplified product. The oligonucleotides of the invention may be used to

CC diagnose HIV-1 infection. The presents sequence represents a capture

CC oligonucleotide probe used to purify RNA in the HIV detection method of

CC the invention.

XX Sequence 57 BP; 35 A; 8 C; 4 G; 10 T; 0 other;

Query Match 2.8%; Score 34; DB 24; Length 57;

Best Local Similarity 100.0%; Pred. No. 0.0011; Mismatches 0; Indels 0; Gaps 0;

Oy 1184 CGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1217  
 Db 23 CGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 56

RESULT 7

AA45493 ID AA45493 standard; DNA; 57 BP.

XX AA45493;



```

XX DT 06-JUN-2002 (first entry)
XX DE HIV-1 RNA capture oligonucleotide SEQ ID NO: 31.
XX KW HIV-1; gag gene; pol gene; PCR; primer; drug resistance; genetic subtype;
XX KW probe; ss.
XX OS Human immunodeficiency virus type 1.
XX PN WO200220852-A1.
XX PD 14-MAR-2002.
XX PF 01-SEP-2000; 2000WO-US24117.
XX PR 01-SEP-2000; 2000WO-US24117.
XX PA (GENP-) GEN-PROBE INC.
XX PA (INMR) BIOMERIEUX SA.
XX PI Yang YY, Brentano ST, Babola O, Tran N, Vernet G;
XX DR WPI; 2002-292273/33.
XX PR New nucleic acid oligomer, useful for detecting selected regions of gag
XX PT and pol genes of human immune deficiency virus, particularly for
XX PT assessing drug resistance -
XX XX
XX XX Example 4; Page 57; 82pp; English.
XX XX
XX CC The present invention provides a number of nucleic acid oligomers which
XX CC can be used to amplify the gag and pol genes of human immunodeficiency
XX CC virus type I (HIV-1). These are used to detect regions of the gag and pol
XX CC genes, especially regions associated with drug resistance, and also for
XX CC identifying genetic subtypes of the virus. The present sequence is an
XX CC oligomer of the invention.
XX XX
XX SQ -Sequence 57 BP; 35 A; 8 C; 4 G; 10 T; 0 other;
Query Match 2.8%; Score 34; DB 24; Length 57;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1184 CGTTTAAAAA...AAAAAAAAAAAAAAAAAAAA 1217
Db 23 CGTTTAAAAA...AAAAAAAAAAAAAAAAAAAA 56

RESULT 8
ABV61939/c
ID ABV61939 standard; cDNA; 310 BP.
XX AC ABV61939;
XX XX
XX DT 13-SEP-2002 (first entry)
XX DE Human prostate expression marker CDNA 61930.
XX XX
XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX KW pharmacogenomic marker; gene; ss.
XX OS Homo sapiens.
XX PN WO200160860-A2.
XX PD 23-AUG-2001.
XX PF 20-FEB-2001; 2001WO-US05171.
XX PR 17-FEB-2000; 2000US-183319P.
XX PR 16-MAR-2000; 2000US-189862P.
XX PR 25-MAY-2000; 2000US-207454P.

```

```

PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX Schlegel R, Endege WO, Monahan JE;
XX WPI; 2001-662795/76.
XX XX
XX PT Novel isolated nucleic acid molecule associated with cancerous state of
XX PT prostate cells and correlating with presence of prostate cancer, useful
XX PT for detecting presence of prostate cancer, stage of prostate cancer -
XX XX
XX PS Claim 1; Page 11710; 11750pp; English.
XX XX
XX CC The invention relates to an isolated nucleic acid molecule (I) comprising
XX CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX CC specification or its complement. (I) is useful for:
XX CC (a) assessing whether a patient is afflicted with prostate cancer;
XX CC (b) monitoring the progression of prostate cancer in a patient;
XX CC (c) assessing the efficacy of a test compound to inhibit prostate
XX CC cancer in a patient;
XX CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
XX CC in a patient;
XX CC (e) selecting a composition for inhibiting prostate cancer in a patient;
XX CC (f) assessing the prostate cell carcinogenic potential of a compound;
XX CC (g) determining whether prostate cancer has metastasized in a patient;
XX CC (h) assessing the aggressiveness or incidence of prostate cancer in a
XX CC patient;
XX CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX SQ Sequence 310 BP; 101 A; 57 C; 38 G; 114 T; 0 other;
Query Match 2.8%; Score 34; DB 23; Length 310;
Best Local Similarity 100.0%; Pred. No. 0.00082;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1184 CGTTTAAAAA...AAAAAAAAAAAAAAAAAAAA 1217
Db 39 CGTTTAAAAA...AAAAAAAAAAAAAAAAAAAA 6

RESULT 9
AAI87403
ID AAI87403 standard; cDNA; 407 BP.
XX AC AAI87403;
XX XX
XX DT 06-NOV-2001 (first entry)
XX DE Human polynucleotide SEQ ID NO 7463.
XX XX
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX KW tissue growth factor; immunomodulatory; cancer; leukaemia;
XX KW nervous system disorders; arthritis; inflammation; ss.
XX OS Homo sapiens.
XX PN WO200164835-A2.
XX PD 07-SEP-2001.
XX PF 26-FEB-2001; 2001WO-US04927.
XX PR 28-FEB-2000; 2000US-0515126.
XX PR 18-MAY-2000; 2000US-0577409.
XX PA (HYSE-) HYSEQ INC.
XX PT Tang YT, Liu C, Drmanac RT;
XX XX

```

DR WPI: 2001-514838/56.  
DR P-PSDB; AA007472.  
PT Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukemia, inflammation and immune  
PT disorders -  
XX  
XX  
XX Claim 1; SEQ ID NO 7463; 1399bp + Sequence Listing; English.  
XX  
CC The invention relates to human polynucleotides (AA179941-AA193841) and  
CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoietic regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 407 BP; 165 A; 65 C; 80 G; 83 T; 14 other;  
SQ  
Query Match 2.8%; Score 34; DB 22; Length 407;  
Best Local Similarity 100.0%; Pred. No. 0.00078;  
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1184 CGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1217  
Db 57 CGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 90  
RESULT 10  
AAC87257  
ID AAC87257 standard; cDNA; 1908 BP.  
AC  
XX AAC87257;  
AC  
DT 09-MAR-2001 (first entry)  
DT  
XX Mouse liver growth hormone-induced cDNA clone 5, SEQ ID NO:7.  
DE  
XX Mouse; growth hormone; GH regulatable gene; liver pathology; hypertrophy;  
KM hepatocellular lesion; hyperplasia; altered expression level; clone 5;  
KM diagnostic marker; gigantism; acromegaly; diabetes; hepatocytotic;  
KM transgenic animal; drug screening; drug discovery; murine; ss.  
XX  
XX Mus sp.  
OS  
XX  
XX MO200066787-A2.  
PN  
XX  
XX 09-NOV-2000.  
PD  
XX 05-MAY-2000; 2000MO-US12366.  
PF  
XX 05-MAY-1999; 99US-0132663.  
PR  
XX (UYOH-) UNIV OHIO.  
PA  
XX  
XX Kogchick JJ, Tjong J;  
PI  
XX  
XX WPI: 2001-007239/01.  
DR P-PSDB; AAB48724, AAB48725, AAB48726, AAB48727, AAB48728, AAB48729,  
DR AAB48730, AAB48731, AAB48732, AAB48733, AAB48734, AAB48735, AAB48736,  
DR AAB48737, AAB48738, AAB48739.  
XX  
XX  
XX Diagnosing abnormal levels of growth hormone activity in liver  
PT comprising assaying growth transcriptional activity and protein  
PT expression level of hormone-regulatable liver genes, as diagnostic  
PT markers of liver pathology -

XX  
XX Claim 1; Page 51-52; 65pp; English.  
XX  
XX  
CC The invention relates to a method of diagnosing abnormal levels of  
CC growth hormone (GH) activity in the liver, or predicting a change in the  
CC condition of the liver in response to abnormal GH activity. The method  
CC involves correlating the level of expression of certain specific genes  
CC with the level of GH activity in the liver, or with an expected change  
CC in the condition of the liver as the result of GH activity. Excessive GH  
CC activity in the liver is thought to be deleterious to health, causing an  
CC increase in liver size as a consequence of both hyperplasia and  
CC hepatocyte hypertrophy, and hepatocellular lesions which progress with  
CC age. Studies in transgenic mice which express high levels of bovine  
CC growth hormone identified a number of genes whose expression in the  
CC liver is altered by high GH levels. The genes which are upregulated are  
CC those encoding alpha-fetoprotein, corticosteroid binding globulin,  
CC fetuin, rab8-interacting protein, paraoxonase-3, cytochrome P45011A,  
CC S-2 hydroxyacid oxidase, interferon alpha/beta receptor, growth hormone  
CC receptor, proteasome 2-subunit, and coagulation factor V. Two novel  
CC genes, clone 5 (AAC87257) and clone 45 (AAC87258) are also upregulated  
CC in response to abnormally high GH levels. Conversely, expression of the  
CC gene encoding 3-beta-hydroxysteroid dehydrogenase/delta-5-delta-4  
CC isomerase appears to be downregulated. The invention also relates to  
CC transgenic nonhuman mammals comprising a GH-induced transgene which  
CC exhibit or have a propensity to develop a liver pathology; an assay for  
CC drugs which inhibit the development of, or which treat a liver pathology,  
CC comprising administering the drug to the transgenic animal; and  
CC administering a drug which inhibits the expression of a GH-induced gene.  
CC The method of the invention is used for diagnosing abnormal levels of GH  
CC activity in the liver or predicting a change in the condition of the  
CC liver in response to abnormal levels of GH activity. The GH-regulatable  
CC liver gene and proteins are useful as diagnostic markers of liver  
CC pathology. Assays for the expression of these genes is useful for the  
CC diagnosis of liver pathologies associated with gigantism or acromegaly or  
CC with diabetes, as other causative agents may act directly or indirectly  
CC upon the same genes. The present sequence represents the novel mouse  
CC liver cDNA clone 5, expression of which is upregulated by abnormal GH  
CC levels.  
XX  
XX  
SQ Sequence 1908 BP; 525 A; 475 C; 477 G; 431 T; 0 other;  
Query Match 2.8%; Score 34; DB 22; Length 1908;  
Best Local Similarity 100.0%; Pred. No. 0.00058;  
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1184 CGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1217  
Db 1814 CGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1847  
RESULT 11  
AAA52773  
ID AAA52773 standard; cDNA; 2040 BP.  
AC  
XX AAA52773;  
AC  
DT 03-JAN-2001 (first entry)  
DT  
XX Soybean putative catabolite repression protein SNF1 coding sequence #3.  
DE  
XX Soybean; carbon catabolite repression; sucrose non-fermenting protein 1;  
KM SNF1; plant growth; ss.  
KM  
XX  
XX Glycine max.  
OS  
XX  
XX Key Location/Qualifiers  
FH 360..1676  
FT /\*tag= a  
FT /product= "SNF1"  
XX  
XX MO200036115-A2.  
PN  
XX

```
PD 22-JUN-2000.
XX
PF 15-DEC-1999; 99WO-US29824.
XX
PR 16-DEC-1998; 98US-0112563.
XX
PA (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI Allen SM, Heppard EP, Sakai H, Weng Z, Helentjaris TG, Macool DU;
PI Miao G;
XX
XX WPI; 2000-431593/37.
DR P-PSDB; AAB03421.
XX
XX New nucleic acids encoding sucrose non-fermenting 1 (SNF1) proteins
PT involved in carbon catabolite repression in plants and seeds, useful
PT for controlling carbon and nitrogen partitioning pathways during plant
PT growth and development -
XX
XX Claim 2; Page 45; 59pp; English.
XX
XX The present sequence is a putative sucrose non-fermenting protein SNF1
CC coding sequence from the soybean. The sequence was isolated by searching
CC a seed cDNA library for sequences similar to those known to encode SNF1
CC in Arabidopsis thaliana, Cucumis sativus, Glycine max and Oryza sativa.
CC The protein is involved in carbon catabolite repression, and so the gene
CC and protein can be used in plants to control the nitrogen and carbon
CC partitioning pathways during plant growth and development. They can also
CC be used to alter the accumulation of carbohydrates, lipids and proteins
CC during plant growth.
XX
XX Sequence 2040 BP; 544 A; 443 C; 555 G; 498 T; 0 other;
SQ
Query Match 2.8%; Score 34; DB 21; Length 2040;
Best Local Similarity 100.0%; Pred. No. 0.00057;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1184 CGTTTAAAAA...AAAAAAAAAAAAAAAAAAAA 1217
DB 2006 CGTTTAAAAA...AAAAAAAAAAAAAAAAAAAA 2039
RESULT 12
ABL33558
ID ABL33558 standard; DNA; 7128 BP.
XX
XX ABL33558;
AC
XX
XX 26-MAR-2002 (first entry)
DT
DE Human immune system associated gene SEQ ID NO: 1531.
XX
XX Human; immune system disease; cytosine methylation; antiasthmatic;
XX antiarteriosclerotic; antianaemic; cytostatic; neotropic;
XX neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
XX antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
XX gene; ds.
XX
XX Homo sapiens.
OS
XX
XX WQ200200928-A2.
PN
XX
XX 03-JAN-2002.
PD
XX
XX 02-JUL-2001; 2001WO-EP07537.
PF
XX
XX 30-JUN-2000; 2000DE-1032529.
PR
XX
XX 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIG-) EPIGENOMICS AG.
PA
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2002-130909/17.
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
XX Claim 1; SEQ ID NO 1531; 32pp + Sequence Listing; German.
XX
XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
XX Sequence 7128 BP; 2031 A; 207 C; 1612 G; 3278 T; 0 other;
SQ
Query Match 2.8%; Score 34; DB 24; Length 7128;
Best Local Similarity 100.0%; Pred. No. 0.00045;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1182 TTCGTTTAAAAA...AAAAAAAAAAAAAAAAAAAA 1215
DB 1188 TTCGTTTAAAAA...AAAAAAAAAAAAAAAAAAAA 1221
RESULT 13
ABQ67149
ID ABQ67149 standard; DNA; 40324 BP.
XX
XX ABQ67149;
AC
XX
XX 28-AUG-2002 (first entry)
DT
DE Human angiogenesis associated polynucleotide SEQ ID NO 179.
XX
XX Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
XX inflammation; rheumatoid arthritis; diabetic retinopathy; antiulcers;
XX macular degeneration; inflammatory bowel disease; Crohn's disease;
XX antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
XX antiarteriosclerotic; ds.
XX
XX Homo sapiens.
OS
XX
XX WQ200246454-A2.
PN
XX
XX 13-JUN-2002.
PD
XX
XX 06-DEC-2001; 2001WO-EP14320.
PF
XX
XX 06-DEC-2000; 2000DE-1061338.
PR
XX
XX (EPIG-) EPIGENOMICS AG.
PA
XX Schacht O;
XX
XX WPI; 2002-500450/53.
XX
XX New nucleic acid fragments from chemically treated
PT angiogenesis-associated genes, useful for determining methylation
PT status, e.g. in diagnosis or treatment of cancer -
XX
XX Claim 1; SEQ ID NO 179; 41pp + Sequence Listing; German.
XX
XX The invention relates to a nucleic acid (I) comprising a segment of 18
CC bases of chemically pretreated DNA of angiogenesis-associated genes (II)
CC having sequences (ABQ66971-ABQ67178) or their complements. (I) also
CC related oligomers, are used to evaluate the methylation status and/or
```

CC single-nucleotide polymorphisms, in angiogenesis-related genes, for  
CC diagnosis and treatment of eye diseases, proliferative retinopathy,  
CC neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,  
CC diabetic retinopathy, macular degeneration caused by neovascularisation,  
CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and  
CC Crohn's disease.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcc\_sequences.  
CC  
XX

SQ Sequence 40324 BP; 12279 A; 436 C; 8127 G; 19482 T; 0 other;

Query Match 2.8%; Score 34; DB 24; Length 40324;

Best Local Similarity 100.0%; Pred. No. 0.00033;

Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1182 TTTCGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1215

DB 1188 TTTCGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1221

RESULT 14

ABK93908

ID ABK93908 standard; DNA; 51 BP.

AC ABK93908;

DT 26-AUG-2002 (first entry)

XX Human immunodeficiency virus type 2 detection probe #32.

XX Human immunodeficiency virus type 2 detection probe #32.

KM Human immunodeficiency virus 2; HIV-2; HIV-1; hepatitis B virus;

KM HBV; hepatitis C virus; HCV; probe: se.

OS Human immunodeficiency virus type 2.

PN WO200234951-A2.

XX 02-MAY-2002.

PF 22-OCT-2001; 2001WO-US45396.

PR 23-OCT-2000; 2000US-242620P.

XX 30-MAR-2001; 2001US-280058P.

PA (GENP-) GEN-PROBE INC.

XX Yang YV, Burrell TA;

XX WPI; 2002-489953/52.

PT Detecting human immunodeficiency virus-2 nucleic acids in a sample, by

PT amplifying nucleic acids with oligonucleotides and detecting nucleic

PT probe:target duplex -

PS Disclosure; Page 57; 58pp; English.

XX The invention relates to a method of detecting human immunodeficiency

XX virus-2 nucleic acids (NA) in a biological sample, by contacting NAs

XX with 1st and 2nd amplification oligonucleotides, amplifying HIV-2

XX nucleic acid base sequence, and detecting the amplified NA; or providing a

XX hybridisation probe and detectable label, hybridising HIV-2 NA with the

XX probe to form probe:target duplex, and detecting the duplex. The

XX method is useful for detecting the presence of HIV-2 NAs in a lysate

XX or a blood product such as plasma or serum, and also for detecting

XX subtypes A, B, C and D of HIV-2. The method is useful for amplifying

XX and detecting the NA in blood serum and also as components of multiple

XX amplification reactions that synthesise amplicons corresponding to

XX polynucleotides of unrelated viruses, e.g. HIV-1, hepatitis B virus

XX (HBV) and hepatitis C virus (HCV). ABK93877-ABK93910 represent

XX HIV-2 detection probes of the invention.

SQ Sequence 51 BP; 31 A; 8 C; 3 G; 9 T; 0 other;

Query Match 2.7%; Score 33; DB 24; Length 51;

Best Local Similarity 100.0%; Pred. No. 0.0028;

Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1185 GTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1217

DB 18 GTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 50

RESULT 15

AAS63020

ID AAS63020 standard; CDNA; 116 BP.

XX AAS63020;

DT 29-JAN-2002 (first entry)

XX Cell death protective sequence CNI-00725, ORF #11.

XX Human; protective sequence; cell death; cerebral oedema; infection;

XX meningitis; degenerative disease; Alzheimer's disease; heart disease;

XX motor neuron disease; demyelinating disease; multiple sclerosis; asthma;

XX nutritional condition; peripheral nervous system disorder; ischaemia;

XX diabetic neuropathy; autoimmune haemolytic anaemia; respiratory system;

XX oral cavity; gastrointestinal tract; liver; cirrhosis; pancreatitis;

XX polycystic renal disease; urinary tract; genitalia; endometriosis;

XX breast; chronic mastitis; thyroid gland; Hashimoto's thyroiditis;

XX adrenal gland; skin; psoriasis; muscular atrophy; bone marrow;

XX osteoporosis; cancer; autoimmune disease; ss.

XX Homo sapiens.

OS Homo sapiens.

PN WO200176532-A2.

XX 18-OCT-2001.

PF 09-APR-2001; 2001WO-US11655.

PR 11-APR-2000; 2000US-0547596.

PA (COGE-) COGENT NEUROSCIENCE INC.

XX Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC, Lo DC;

XX WPI; 2002-017408/02.

XX P-PSDB; AAU69621.

PT Novel nucleic acids referred as protective sequences and their encoded

PT products for diagnosing, treating diseases involving cell death,

PT including neurological disorders e.g. stroke and for identifying

PT modulators -

PS Claim 2; Figure 6K; 256pp; English.

XX The invention relates to isolated protective sequence polypeptides (I)

XX and polynucleotides (II). (II) is useful for transferring a protective

XX sequence into a cell, which delays and/or prevents the cell from

XX undergoing cell death. Protective sequences, their products or

XX antibodies are useful diagnostically, prophylactically, therapeutically

XX or as targets for treatment and diagnosis of conditions, disorders or

XX diseases involving cell death. The protective sequences and their

XX products are useful for preventing or treating disorders of the central

XX nervous system including neurological and psychiatric conditions,

XX cerebral oedema, infections such as meningitis, degenerative diseases

XX such as Alzheimer's and motor neuron disease, demyelinating diseases such

XX as multiple sclerosis, nutritional conditions, disorders of the

XX peripheral nervous system including diabetic neuropathy, disorders

XX which cause cell death in organ systems including blood vessels, heart

XX (ischaemia), blood cells (autoimmune haemolytic anaemia), respiratory

XX system (asthma), oral cavity, gastrointestinal tract, liver (cirrhosis),

XX pancreatitis, polycystic renal disease, urinary tract, genitalia

CC (congenital anomalies), endometriosis, breast (chronic mastitis), thyroid  
CC gland (Hashimoto's thyroiditis), adrenal gland, skin (psoriasis),  
CC musculoskeletal system (muscular atrophy), bone marrow or bone  
CC (osteoporosis). The compositions promote cell death and are useful for  
CC treating and/or ameliorating cancer and autoimmune diseases. The  
CC compounds are further useful for treating physiological impacts on  
CC organs caused by infection which induce cell death. (i) is useful to  
CC raise an immune response, as a reagent in assays designed to  
CC quantitatively determine levels of the protein in biological fluids, as  
CC markers for tissues in which the corresponding protein is expressed and  
CC to isolate receptors or ligands. AAS62967-AAS63142 represent the  
CC protective polynucleotide sequences as described in the invention.

XX  
SQ Sequence 116 BP; 73 A; 18 C; 11 G; 14 T; 0 other;

Query Match 2.7%; Score 33; DB 24; Length 116;  
Best Local Similarity 100.0%; Pred. No. 0.0024;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1185 GTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1217

Db 50 GTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 82

Search completed: December 17, 2002, 17:47:35  
Job time : 310 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 17, 2002, 17:10:36 ; Search time 74 Seconds  
(without alignments)  
5043.589 Million cell updates/sec

Title: US-09-720-384A-3

Perfect score: 1217  
Sequence: 1 gcgtccgtttcattcatca.....aaaaaaaaaaaaaaaaaa 1217

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PCTUS COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	2.8	57	US-09-620-958A-7	Sequence 7, Appli
2	33	2.7	1092	US-09-227-357-35	Sequence 35, Appli
3	33	2.7	1738	US-08-373-482A-2	Sequence 2, Appli
4	33	2.7	1910	US-09-009-438-1	Sequence 1, Appli
5	33	2.7	1910	US-09-207-493-1	Sequence 1, Appli
6	33	2.7	1951	US-09-465-558-35	Sequence 35, Appli
7	33	2.7	2000	US-08-351-473B-1	Sequence 1, Appli
8	33	2.7	2790	US-08-800-291B-1	Sequence 1, Appli
9	33	2.7	3159	US-09-437-054A-7	Sequence 7, Appli
10	33	2.7	3410	US-09-020-956-110	Sequence 110, App
11	33	2.7	3410	US-09-030-607-110	Sequence 110, App
12	33	2.7	3410	US-09-605-785-110	Sequence 110, App
13	33	2.7	3410	US-09-439-313-110	Sequence 110, App
14	33	2.7	3410	US-09-352-616A-110	Sequence 110, App
15	33	2.7	3410	US-09-602-877A-100	Sequence 100, App
16	33	2.7	3410	US-09-232-149A-110	Sequence 110, App
17	33	2.7	3602	US-09-402-929-1	Sequence 1, Appli
18	32	2.6	67	US-09-620-958A-6	Sequence 6, Appli
19	32	2.6	114	US-08-120-817-99	Sequence 99, Appli
20	32	2.6	114	US-08-478-675-99	Sequence 99, Appli
21	32	2.6	144	US-08-702-344-26	Sequence 26, Appli
22	32	2.6	222	US-08-481-190-15	Sequence 15, Appli
23	32	2.6	222	PCT-US93-00869-15	Sequence 15, Appli
24	32	2.6	375	US-08-946-026-23	Sequence 23, Appli
25	32	2.6	399	US-07-885-970A-13	Sequence 13, Appli
26	32	2.6	399	US-08-298-687A-13	Sequence 13, Appli
27	32	2.6	399	US-08-530-797-12	Sequence 12, Appli

28	32	2.6	399	1	US-08-298-829-13	Sequence 13, Appli
29	32	2.6	399	2	US-08-787-335-12	Sequence 12, Appli
30	32	2.6	522	2	US-08-909-965C-16	Sequence 16, Appli
31	32	2.6	724	4	US-09-020-956-38	Sequence 38, Appli
32	32	2.6	724	4	US-09-030-607-38	Sequence 38, Appli
33	32	2.6	724	4	US-09-605-785-38	Sequence 38, Appli
34	32	2.6	724	4	US-09-439-313-38	Sequence 38, Appli
35	32	2.6	724	4	US-09-352-616A-38	Sequence 38, Appli
36	32	2.6	724	4	US-09-232-149A-38	Sequence 38, Appli
37	32	2.6	911	2	US-08-924-759-9	Sequence 9, Appli
38	32	2.6	911	3	US-09-248-335-9	Sequence 9, Appli
39	32	2.6	966	1	US-08-514-014-7	Sequence 7, Appli
40	32	2.6	966	2	US-08-833-823-7	Sequence 7, Appli
41	32	2.6	976	2	US-08-504-459-9	Sequence 9, Appli
42	32	2.6	980	4	US-09-452-239-21	Sequence 21, Appli
43	32	2.6	991	3	US-08-924-747-25	Sequence 25, Appli
44	32	2.6	991	4	US-09-247-373B-25	Sequence 25, Appli
45	32	2.6	991	4	US-09-296-715-25	Sequence 25, Appli

ALIGNMENTS

RESULT 1  
US-09-620-958A-7  
; Sequence 7, Application US/09620958A  
; Patent No. 6294338  
; GENERAL INFORMATION:  
; APPLICANT: Nuchimura, Kiyorada  
; TITLE OF INVENTION: POLYNUCLEOTIDE AMPLIFICATION METHOD  
; FILE REFERENCE: GP104-02.UT  
; CURRENT APPLICATION NUMBER: US/09/620,958A  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 7  
; LENGTH: 57  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Sequence of the (-)4258 A30 capture probe.  
; Patent No. 6294338  
US-09-620-958A-7

Query Match 2.8%; Score 34; DB 4; Length 57;  
Best Local Similarity 100.0%; Pred. No. 1e-05; 0;  
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1184 CGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1217  
Db 23 CGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 56

RESULT 2  
US-09-227-357-35  
; Sequence 35, Application US/09227357  
; Patent No. 6342581  
; GENERAL INFORMATION:  
; APPLICANT: Fischer et al.  
; TITLE OF INVENTION: 123 Human Secreted Proteins  
; FILE REFERENCE: P2010P1  
; CURRENT APPLICATION NUMBER: US/09/227,357  
; CURRENT FILING DATE: 1999-01-08  
; EARLIER APPLICATION NUMBER: PCT/US98/13684  
; EARLIER FILING DATE: 1998-07-07  
; EARLIER APPLICATION NUMBER: 60/051,926  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,793  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,925  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,929  
; EARLIER FILING DATE: 1997-07-08

EARLIER APPLICATION NUMBER: 60/052,803  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/052,732  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,931  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,932  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,916  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,930  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,918  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,920  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/052,733  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/052,795  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,919  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,928  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/055,722  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,723  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,948  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,949  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,953  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,950  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,947  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,964  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/056,360  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,684  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,984  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,954  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/058,785  
EARLIER FILING DATE: 1997-09-12  
EARLIER APPLICATION NUMBER: 60/058,664  
EARLIER FILING DATE: 1997-09-12  
EARLIER APPLICATION NUMBER: 60/058,660  
EARLIER FILING DATE: 1997-09-12  
EARLIER APPLICATION NUMBER: 60/058,661  
EARLIER FILING DATE: 1997-09-12  
NUMBER OF SEQ ID NOS: 672  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 35  
LENGTH: 1092  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-227-357-35

Query Match 2.7%; Score 33; DB 4; Length 1092;  
Best Local Similarity 100.0%; Pred. No. 2e-05;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1185 GTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1217  
DB 1053 GTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1085

RESULT 3  
US-08-379-482A-2  
Sequence 2, Application US/08379482A  
Patent No. 5859334  
GENERAL INFORMATION:  
APPLICANT: Brugliera, Filippa  
APPLICANT: Holton, Timothy A.  
TITLE OF INVENTION: GENETIC SEQUENCES ENCODING  
TITLE OF INVENTION: GYCOSYLTRANSFERASE ENZYMES AND USES THEREFOR  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Scully, Scott, Murphy & Presser  
STREET: 400 Garden City Plaza  
CITY: Garden City  
STATE: New York  
COUNTRY: USA  
ZIP: 11530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/379,482A  
FILING DATE: 30-JUL-1993  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Digilio, Frank S.  
REGISTRATION NUMBER: 31,346  
REFERENCE/DOCKET NUMBER: 9590  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (516)742-4343  
TELEFAX: (516)742-4366  
TELEX: 230 901 SANS UR  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1738 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1413  
US-08-379-482A-2  
Query Match 2.7%; Score 33; DB 2; Length 1738;  
Best Local Similarity 100.0%; Pred. No. 1.9e-05;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 1185 GTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1217  
DB 1667 GTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1699  
RESULT 4  
US-09-009-438-1  
Sequence 1, Application US/09009438  
Patent No. 5981223  
GENERAL INFORMATION:  
APPLICANT: SATHE, GANESH M.  
APPLICANT: HALSEY, WENDY S.  
APPLICANT: BERGSM, DEBK J.  
TITLE OF INVENTION: HDK017 ENCODES A NOVEL 7-TRANSMEMBRANE  
TITLE OF INVENTION: RECEPTOR  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: RATNER & PRESTIA  
STREET: P. O. BOX 980  
CITY: VALLEY FORGE  
STATE: PA  
COUNTRY: USA  
ZIP: 19482



COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/009,438  
FILING DATE: 20-JAN-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: PRESTIA, PAUL F  
REGISTRATION NUMBER: 23,031  
REFERENCE/DOCKET NUMBER: GH-70369  
TELEPHONE: 610-407-0700  
TELEFAX: 610-407-0701  
TELEX: 846169  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1910 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-09-009-438-1

Query Match 2.7%; Score 33; DB 2; Length 1910;  
Best Local Similarity 100.0%; Pred. No. 1.9e-05;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1185 GTTTAAAAA 1217  
|||||  
DB 1859 GTTTAAAAA 1891

## RESULT 5

US-09-207-493-1  
Sequence 1, Application US/09207493  
Patent No. 6114127  
GENERAL INFORMATION:  
APPLICANT: Sathe, Gimesh M.  
APPLICANT: Halsey, Wendy S.  
APPLICANT: Bergsma, Derek J.  
APPLICANT: Culp, Jeffrey S.  
APPLICANT: Wang, Da-Yuan  
TITLE OF INVENTION: METHODS OF SCREENING FOR AGONISTS AND  
TITLE OF INVENTION: ANTAGONISTS OF THE HDPXU17 RECEPTOR  
FILE REFERENCE: GH70369-1  
CURRENT APPLICATION NUMBER: US/09/207,493  
CURRENT FILING DATE: 1998-12-08  
EARLIER APPLICATION NUMBER: 09/009,438  
EARLIER FILING DATE: 1998-01-20  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 1910  
TYPE: DNA  
ORGANISM: Human  
US-09-207-493-1

Query Match 2.7%; Score 33; DB 3; Length 1910;  
Best Local Similarity 100.0%; Pred. No. 1.9e-05;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1185 GTTTAAAAA 1217  
|||||  
DB 1859 GTTTAAAAA 1891

## RESULT 6

US-09-465-558-35  
Sequence 35, Application US/09465558  
Patent No. 6436657  
GENERAL INFORMATION:  
APPLICANT: Morakinyo, Layo O.  
APPLICANT: Orozco Jr., Emil M.  
TITLE OF INVENTION: TETRAHYDROFOLATE METABOLIC ENZYMES  
FILE REFERENCE: BB1322 US NA  
CURRENT APPLICATION NUMBER: US/09/465,558  
CURRENT FILING DATE: 1999-12-17  
EARLIER APPLICATION NUMBER: 60/112,734  
EARLIER FILING DATE: 1998-12-18  
NUMBER OF SEQ ID NOS: 70  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 35  
LENGTH: 1951  
TYPE: DNA  
ORGANISM: Oryza sativa  
US-09-465-558-35

Query Match 2.7%; Score 33; DB 4; Length 1951;  
Best Local Similarity 100.0%; Pred. No. 1.9e-05;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1185 GTTTAAAAA 1217  
|||||  
DB 1900 GTTTAAAAA 1932

## RESULT 7

US-08-351-473B-1  
Sequence 1, Application US/08351473B  
Patent No. 5656440  
GENERAL INFORMATION:  
APPLICANT: LENZEN, GERLINDA  
APPLICANT: KAPOOR, ARCHANA  
TITLE OF INVENTION: NUCLEOTIDE SEQUENCES CODING FOR THE  
TITLE OF INVENTION: BOVINE BETAS-ADRENERGIC RECEPTOR AND THEIR APPLICATIONS  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT  
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
CITY: ARLINGTON  
STATE: VIRGINIA  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/351,473B  
FILING DATE: 21-FEB-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 93 04670  
FILING DATE: 21-APR-1993  
PRIOR APPLICATION DATA: PCT/FR94/00447  
APPLICATION NUMBER: PCT/FR94/00447  
FILING DATE: 21-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: OBLON, NORMAN F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 6639-001-0X PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2000 base pairs  
TYPE: nucleic acid

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 107..1321  
OTHER INFORMATION: /function="BOVINE BETA-3 RECEPTOR"  
OTHER INFORMATION: /product="ADRENERGIC, BETA RECEPTOR"  
US-08-351-473B-1

Query Match 2.7%; Score 33; DB 1; Length 2000;  
Best Local Similarity 100.0%; Pred. No. 1.9e-05;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1185 GTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1217  
Db 1956 GTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1988

RESULT 8  
US-08-800-291B-1  
Sequence 1, Application US/08800291B  
Patent No. 6153740  
GENERAL INFORMATION:  
APPLICANT: J.D. Young & C.E. Cass  
TITLE OF INVENTION: CDNA ENCODING NUCLEOSIDE TRANSPORTER  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/800,291B  
FILING DATE: 13-FEB-1997  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/499,314  
FILING DATE: 7-JULY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile, Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 07254/044WO1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/678-5099  
TELEFAX: 619/678-5070  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2790 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..2790  
US-08-800-291B-1

Query Match 2.7%; Score 33; DB 3; Length 2790;  
Best Local Similarity 100.0%; Pred. No. 1.9e-05;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Cy 1185 GTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1217  
Db 2729 GTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2761

RESULT 9  
US-09-437-054A-7  
Sequence 7, Application US/09437054A  
Patent No. 6316698  
GENERAL INFORMATION:  
APPLICANT: Allen, Stephen M.  
APPLICANT: Kinney, Anthony J.  
TITLE OF INVENTION: Plant Alpha-Glucosidase II Homologs  
FILE REFERENCE: BB1273 US NA  
CURRENT APPLICATION NUMBER: US/09/437,054A  
CURRENT FILING DATE: 2001-05-14  
PRIOR APPLICATION NUMBER: 60/107,909  
PRIOR FILING DATE: 1998-No. 6316698ember-10  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 7  
LENGTH: 3159  
TYPE: DNA  
ORGANISM: Glycine max  
US-09-437-054A-7

Query Match 2.7%; Score 33; DB 4; Length 3159;  
Best Local Similarity 100.0%; Pred. No. 1.8e-05;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1185 GTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1217  
Db 3116 GTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3148

RESULT 10  
US-09-020-956-110  
Sequence 110, Application US/09020956  
Patent No. 6261562  
GENERAL INFORMATION:  
APPLICANT: Xu, Jiangchun  
APPLICANT: Dillin, David C.  
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO  
NUMBER OF SEQUENCES: 178  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED AND BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/020,956  
FILING DATE: 09-FEB-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Makl, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.427C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 110:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3410 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
US-09-020-956-110

```
Query Match          2.7%; Score 33; DB 4; Length 3410;
Best Local Similarity 100.0%; Pred. No. 1.8e-05;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1185 GTTTAAAAA 1217
Db 3322 GTTTAAAAA 3354

RESULT 11
US-09-030-607-110
; Sequence 110, Application US/09030607
; Patent No. 6262245
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
; NUMBER OF SEQUENCES: 224
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; APPLICATION DATA:
; APPLICATION NUMBER: US/09/030,607
; FILING DATE: 25-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.427C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 110:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3410 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; US-09-030-607-110

Query Match          2.7%; Score 33; DB 4; Length 3410;
Best Local Similarity 100.0%; Pred. No. 1.8e-05;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1185 GTTTAAAAA 1217
Db 3322 GTTTAAAAA 3354

RESULT 12
US-09-605-785-110
; Sequence 110, Application US/09605785
; Patent No. 6321716
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
```

```
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C16
; CURRENT APPLICATION NUMBER: US/09/605,785
; CURRENT FILING DATE: 2000-06-27
; NUMBER OF SEQ ID NOS: 835
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 110
; LENGTH: 3410
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-605-785-110

Query Match          2.7%; Score 33; DB 4; Length 3410;
Best Local Similarity 100.0%; Pred. No. 1.8e-05;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1185 GTTTAAAAA 1217
Db 3322 GTTTAAAAA 3354

RESULT 13
US-09-439-313-110
; Sequence 110, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqi
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 110
; LENGTH: 3410
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-439-313-110

Query Match          2.7%; Score 33; DB 4; Length 3410;
Best Local Similarity 100.0%; Pred. No. 1.8e-05;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1185 GTTTAAAAA 1217
Db 3322 GTTTAAAAA 3354

RESULT 14
US-09-352-616A-110
; Sequence 110, Application US/09352616A
```

```

; Patent No. 6395278
; GENERAL INFORMATION:
; APPLICANT: Dillon, Davin C.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang, Yugu
; APPLICANT: Xu, Jianshun
; APPLICANT: Mitcham, Jennifer Lynn
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.427C8
; CURRENT APPLICATION NUMBER: US/09/352,616A
; CURRENT FILING DATE: 1999-07-13
; NUMBER OF SEQ ID NOS: 472
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO. 110
; LENGTH: 3410
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-352-616A-110

```

```

Query Match          2.7%: Score 33; DB 4; Length 3410;
Best Local Similarity 100.0%; Pred. No. 1.8e-05;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1185 GTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1217
DB 3322 GTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3354

```

```

RESULT 15
US-09-602-877A-100
; Sequence 100, Application US/09602877A
; Patent No. 6432707
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Xu, Jianshun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.446C5
; CURRENT APPLICATION NUMBER: US/09/602,877A
; CURRENT FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO. 100
; LENGTH: 3410
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-602-877A-100

```

```

Query Match          2.7%: Score 33; DB 4; Length 3410;
Best Local Similarity 100.0%; Pred. No. 1.8e-05;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1185 GTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1217
DB 3322 GTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3354

```

```

Search completed: December 17, 2002, 19:24:49
Job time : 105 secs

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 17, 2002, 17:42:26 ; Search time 80 Seconds  
(without alignments)  
6015.951 Million cell updates/sec

Title: US-09-720-384A-3

Perfect score: 1217

Sequence: 1 ggcgcgtttcttcattcatca.....aaaaaaaaaaaaaaaaaaaa 1217

Scoring table: OLIGO NUC

Gapop 60.0 , Gapext 60.0

Searched: 355320 seqs, 197730502 residues

Word size : 0

Total number of hits satisfying chosen parameters: 710640

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	175	14.4	262	10	US-09-923-876-1680
2	36	3.0	2952	10	US-09-925-297-257
3	34	2.8	57	10	US-09-944-036-31
4	34	2.8	57	10	US-09-943-286-7
5	33	2.7	51	9	US-10-001-407-32
6	33	2.7	178	10	US-09-867-701-9612
c 7	33	2.7	304	10	US-09-867-701-10381
c 8	33	2.7	356	10	US-09-770-791-608
c 9	33	2.7	362	10	US-09-960-352-5390
c 10	33	2.7	376	10	US-09-770-791-242
c 11	33	2.7	413	10	US-09-867-701-243
c 12	33	2.7	418	10	US-09-924-035A-899
c 13	33	2.7	609	10	US-09-770-149-886
c 14	33	2.7	659	10	US-09-770-149-885
15	33	2.7	921	10	US-09-925-297-353
16	33	2.7	1139	10	US-09-764-864-259
17	33	2.7	1143	9	US-09-931-457A-58
c 18	33	2.7	1223	10	US-09-822-830A-204
19	33	2.7	1377	10	US-09-822-830A-588

20	33	2.7	1486	10	US-09-443-704-15	Sequence 15, Appl
21	33	2.7	1531	10	US-09-925-300-577	Sequence 577, Appl
22	33	2.7	1758	10	US-09-764-853-247	Sequence 247, Appl
23	33	2.7	2981	10	US-09-925-301-101	Sequence 101, Appl
24	33	2.7	3410	9	US-09-232-880-110	Sequence 110, Appl
25	33	2.7	3410	9	US-10-012-896-110	Sequence 110, Appl
26	33	2.7	3410	10	US-09-745-288-100	Sequence 100, Appl
27	33	2.7	3410	10	US-09-759-143-110	Sequence 110, Appl
28	33	2.7	3410	10	US-09-780-669-110	Sequence 110, Appl
29	33	2.7	3410	10	US-09-030-605-110	Sequence 110, Appl
30	33	2.7	3410	10	US-09-822-827-110	Sequence 110, Appl
31	33	2.7	3410	10	US-09-115-453-110	Sequence 110, Appl
32	33	2.7	3947	10	US-09-925-300-638	Sequence 638, Appl
33	33	2.7	4640	12	US-10-052-586-75	Sequence 75, Appl
34	33	2.7	4649	9	US-09-978-295A-113	Sequence 113, Appl
35	33	2.7	4649	9	US-09-978-697-113	Sequence 113, Appl
36	33	2.7	4649	9	US-09-978-192A-113	Sequence 113, Appl
37	33	2.7	5520	9	US-10-001-887-43	Sequence 43, Appl
38	33	2.7	326014	10	US-09-731-231A-3	Sequence 3, Appl
39	32	2.6	53	9	US-10-001-407-31	Sequence 31, Appl
40	32	2.6	55	10	US-09-944-036-32	Sequence 32, Appl
41	32	2.6	56	10	US-09-944-036-30	Sequence 30, Appl
c 42	32	2.6	60	9	US-10-057-940-14	Sequence 14, Appl
43	32	2.6	64	9	US-10-057-940-1	Sequence 1, Appl
c 44	32	2.6	64	9	US-10-057-940-1	Sequence 1, Appl
c 45	32	2.6	64	10	US-09-919-580-637	Sequence 637, Appl

#### ALIGNMENTS

#### RESULT 1

US-09-923-876-1680  
; Sequence 1680, Application US/09923876  
; Patent No. US20020013958A1  
; GENERAL INFORMATION:  
; APPLICANT: Lalquadi, Raghunath V.  
; APPLICANT: Kamigaki, Laura V. (Ito)  
; APPLICANT: Sherman, Bradley K.  
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING  
; FILE REFERENCE: PU-0012-1 CON  
; CURRENT APPLICATION NUMBER: US/09/923,876  
; CURRENT FILING DATE: 2001-08-06  
; PRIOR APPLICATION NUMBER: 09/298,329  
; PRIOR FILING DATE: 1999-04-21  
; PRIOR APPLICATION NUMBER: 60/085,331  
; PRIOR FILING DATE: 1998-05-05  
; NUMBER OF SEQ ID NOS: 6332  
; SOFTWARE: PERL Program  
; SEQ ID NO 1680  
; LENGTH: 262  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20020013958A1 700159305H1  
US-09-923-876-1680

Query Match	14.4%;	Score 175;	DB 10;	Length 262;
Best Local Similarity	99.6%;	Pred. No. 1.4e-55;		
Matches 225;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
Qy 705	CTGATGCTGGTGCATATGCAITGTGATCTTCATATCTCCATACAGAGAGATCGTGATG	764		
Db 37	CTGATGCTGGTGCATATGCAITGTGATCTTCATATCTCCATACAGAGAGATCGTGATG	96		
Qy 765	CATGCCGCTCTACTTCCACATTTCTAACTTTATTGAAGTATTATTGATTTGCCCTAA	824		
Db 97	CATGCCGCTCTACTTCCACATTTCTAACTTTATTGAAGTATTATTGATTTGCCCTAA	156		
Qy 825	AAATTTGTAAGCTCGTGATCTCTAAAGGCTTACAGCTTGCACACAGAAAGATTA	884		
Db 157	AAATTTGTAAGCTCGTGATCTCTAAAGGCTTGCACACAGCTTGCACACAGAAAGATTA	216		

QY 885 AAGGTTTCACGTGAATTGATCATACACCAACCAATTAATG 930  
|||||  
DB 217 AAGGTTTCACGTGAATTGATCATACACCAACCAATTAATG 262  
|||||

RESULT 2  
US-09-925-297-257  
Sequence 257, Application US/09925297  
Patent No. US2002081659A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA105  
CURRENT APPLICATION NUMBER: US/09/925,297  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05989  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 928  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 257  
LENGTH: 2952  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-925-297-257

Query Match  
Best Local Similarity 100.0%; Pred. No. 7.1e-05;  
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1182 TTGCTTTAAAAA 1217  
|||||  
DB 2916 TTGCTTTAAAAA 2951  
|||||

RESULT 3  
US-09-944-036-31  
Sequence 31, Application US/09944036  
Patent No. US2002005095A1  
GENERAL INFORMATION:  
APPLICANT: YANG, Yeasing Y.  
APPLICANT: BREYANO, Steven T.  
APPLICANT: BABOLA, Odile  
APPLICANT: TRAV, Nathalie  
APPLICANT: VERNET, Guy  
TITLE OF INVENTION: APPLICATION OF HIV-1 SEQUENCES FOR DETECTION OF  
TITLE OF INVENTION: SEQUENCES ASSOCIATED WITH DRUG-RESISTANCE MUTATIONS  
FILE REFERENCE: GPl14-02.UT  
CURRENT APPLICATION NUMBER: US/09/944,036  
CURRENT FILING DATE: 2001-08-31  
PRIOR APPLICATION NUMBER: US 60/229,790  
PRIOR FILING DATE: 2000-09-01  
NUMBER OF SEQ ID NOS: 70  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 31  
LENGTH: 57  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Capture  
OTHER INFORMATION: oligonucleotide  
US-09-944-036-31

Query Match  
Best Local Similarity 100.0%; Pred. No. 0.0006;  
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1184 CGTTTAAAAA 1217  
|||||  
DB 23 CGTTTAAAAA 56  
|||||

RESULT 4  
US-09-943-286-7  
Sequence 7, Application US/09943286  
Patent No. US2002010668A1  
GENERAL INFORMATION:  
APPLICANT: Nunomura, Kiyotada  
TITLE OF INVENTION: POLYNUCLEOTIDE AMPLIFICATION METHOD  
FILE REFERENCE: GPl04-02.UT  
CURRENT APPLICATION NUMBER: US/09/943,286  
CURRENT FILING DATE: 2001-08-30  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 7  
LENGTH: 57  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Sequence of the (-)4258 A30 capture probe.  
US-09-943-286-7

Query Match  
Best Local Similarity 100.0%; Pred. No. 0.0008;  
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1184 CGTTTAAAAA 1217  
|||||  
DB 23 CGTTTAAAAA 56  
|||||

RESULT 5  
US-10-001-407-32  
Sequence 32, Application US/10001407  
Patent No. US2002017127A1  
GENERAL INFORMATION:  
APPLICANT: Yang, Yeasing  
APPLICANT: Burrell, Terrie  
TITLE OF INVENTION: Compositions and Methods for Detecting  
TITLE OF INVENTION: Human Immunodeficiency Virus 2 (HIV-2)  
FILE REFERENCE: GPl17-03.UT  
CURRENT APPLICATION NUMBER: US/10/001,407  
CURRENT FILING DATE: 2001-10-22  
PRIOR APPLICATION NUMBER: 60/242,620  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/280,058  
PRIOR FILING DATE: 2001-03-30  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 32  
LENGTH: 51  
TYPE: DNA  
ORGANISM: HIV-2  
US-10-001-407-32

Query Match  
Best Local Similarity 100.0%; Pred. No. 0.0019;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1185 GTTTAAAAA 1217  
|||||  
DB 18 GTTTAAAAA 50  
|||||

RESULT 6  
US-09-867-701-9612  
Sequence 9612, Application US/09867701  
Patent No. US2002013237A1  
GENERAL INFORMATION:  
APPLICANT: Aglate, Paul A.  
APPLICANT: Jones, Robert  
APPLICANT: Harlocker, Susan L.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER  
US-09-867-701-9612

Query Match  
Best Local Similarity 100.0%; Pred. No. 0.0019;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9612
; LENGTH: 178
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9612
```

```
Query Match          2.7%; Score 33; DB 10; Length 178;
Best Local Similarity 100.0%; Pred. No. 0.0015;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1185 GTTTAAAAA 1217
|||||
Db 90 GTTTAAAAA 122
```

```
RESULT 7
US-09-867-701-10381/c
; Sequence 10381, Application US/09867701
; Patent No. US2002013237A1
; GENERAL INFORMATION:
; APPLICANT: Agilate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10381
; LENGTH: 304
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-10381
```

```
Query Match          2.7%; Score 33; DB 10; Length 304;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1185 GTTTAAAAA 1217
|||||
Db 44 GTTTAAAAA 12
```

```
RESULT 8
US-09-770-791-608/c
; Sequence 608, Application US/09770791
; Patent No. US20020062014A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jorn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krieker, Waja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
```

```
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; TITLE OF INVENTION: thaliana
; FILE REFERENCE: 2029 (PARA-018PRV)
; CURRENT APPLICATION NUMBER: US/09/770,791
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/178,480
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 608
; LENGTH: 356
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-791-608
```

```
Query Match          2.7%; Score 33; DB 10; Length 356;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1185 GTTTAAAAA 1217
|||||
Db 35 GTTTAAAAA 3
```

```
RESULT 9
US-09-960-352-5390/c
; Sequence 5390, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 5390
; LENGTH: 362
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 23-LIB34-026-Q1-B1-F3
US-09-960-352-5390
```

```
Query Match          2.7%; Score 33; DB 10; Length 362;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1185 GTTTAAAAA 1217
|||||
Db 106 GTTTAAAAA 74
```

```
RESULT 10
US-09-770-791-242/c
; Sequence 242, Application US/09770791
; Patent No. US20020062014A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jorn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
```

```

; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krickler, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hubban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; FILE REFERENCE: 2029 (PARA-018PRV)
; CURRENT APPLICATION NUMBER: US/09/770,791
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/178,480
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 242
; LENGTH: 376
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-791-242
```

```

Query Match          2.7%; Score 33; DB 10; Length 376;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

OY 1185 GTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1217
Db 39 GTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 7
```

```

RESULT 11
US-09-867-701-243/C
; Sequence 243 Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglase, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 243
; LENGTH: 413
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)-(413)
; OTHER INFORMATION: n = A,T,C or G
US-09-867-701-243
```

```

Query Match          2.7%; Score 33; DB 10; Length 413;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

OY 1185 GTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1217
Db 60 GTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 28
```

```

RESULT 12
US-09-924-035A-899/C
; Sequence 899, Application US/09924035A
; Patent No. US20020142319A1
; GENERAL INFORMATION:
; APPLICANT: Glach, Jim
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; TITLE OF INVENTION: thaliana
```

```

; FILE REFERENCE: 2011US
; CURRENT APPLICATION NUMBER: US/09/924,035A
; CURRENT FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: US 60/148,784
; PRIOR FILING DATE: 1999-08-13
; NUMBER OF SEQ ID NOS: 900
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 899
; LENGTH: 418
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-924-035A-899
```

```

Query Match          2.7%; Score 33; DB 10; Length 418;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

OY 1185 GTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1217
Db 56 GTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 24
```

```

RESULT 13
US-09-770-149-886/C
; Sequence 886, Application US/09770149
; Patent No. US20020059663A1
; GENERAL INFORMATION:
; APPLICANT: Goflach, Jörn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Mathew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krickler, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hubban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; FILE REFERENCE: 2024 (PARA-013PRV)
; CURRENT APPLICATION NUMBER: US/09/770,149
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/178,506
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 886
; LENGTH: 609
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-149-886
```

```

Query Match          2.7%; Score 33; DB 10; Length 609;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

OY 1185 GTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1217
Db 61 GTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 29
```

```

RESULT 14
US-09-770-149-485
; Sequence 485, Application US/09770149
```



; Patent No. US20020059663A1  
; GENERAL INFORMATION:  
; APPLICANT: Goriach, Jorn  
; APPLICANT: An, Yong-Qiang  
; APPLICANT: Hamilton, Carol M.  
; APPLICANT: Price, Jennifer L.  
; APPLICANT: Raines, Tracy M.  
; APPLICANT: Yu, Yang  
; APPLICANT: Rameaka, Joshua G.  
; APPLICANT: Page, Amy  
; APPLICANT: Matthew, Abraham V.  
; APPLICANT: Ledford, Brooke L.  
; APPLICANT: Woessner, Jeffrey P.  
; APPLICANT: Haas, William David  
; APPLICANT: Garcia, Carlos A.  
; APPLICANT: Krickler, Maja  
; APPLICANT: Slader, Ted  
; APPLICANT: Davis, Keith R.  
; APPLICANT: Allen, Keith  
; APPLICANT: Hoffman, Neil  
; APPLICANT: Hurban, Patrick  
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis  
; FILE REFERENCE: 2024 (PARA-013PRV)  
; CURRENT APPLICATION NUMBER: US/09/770,149  
; CURRENT FILING DATE: 2001-01-26  
; PRIOR APPLICATION NUMBER: 60/178,506  
; PRIOR FILING DATE: 2000-01-27  
; NUMBER OF SEQ ID NOS: 999  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 485  
; LENGTH: 659  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
; NAME/KEY: misc feature  
; LOCATION: (1)-(659)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-770-149-485

Query Match 2.7%; Score 33; DB 10; Length 659;  
Best Local Similarity 100.0%; Pred. No. 0.0012;  
Matches - 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1185 GTTTAAAAA 1217  
Db 624 GTTTAAAAA 656

RESULT 15  
US-09-925-297-353  
; Sequence 353, Application US/09925297  
; Patent No. US20020081659A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA105  
; CURRENT APPLICATION NUMBER: US/09/925,297  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05989  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 928  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 353  
; LENGTH: 921  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-297-353

Query Match 2.7%; Score 33; DB 10; Length 921;  
Best Local Similarity 100.0%; Pred. No. 0.0011;

Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1185 GTTTAAAAA 1217  
Db 836 GTTTAAAAA 868  
Search completed: December 17, 2002, 19:25:59  
Job time : 87 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 17, 2002, 15:59:41 ; Search time 2404 Seconds  
(without alignments)  
8198.797 Million cell updates/sec

Title: US-09-720-384A-3  
Perfect score: 1217  
Sequence: 1 ggcgtccgtttcatttcattca.....aaaaaaaaaaaaaaaaaaaa 1217

Scoring table: OLIGO NUC  
Gapop 50.0 , Gapext 50.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- EST:\*
- 1: em\_estba:\*
  - 2: em\_esthum:\*
  - 3: em\_estin:\*
  - 4: em\_estov:\*
  - 5: em\_estov:\*
  - 6: em\_estpl:\*
  - 7: em\_estro:\*
  - 8: em\_hic:\*
  - 9: gb\_est1:\*
  - 10: gb\_est2:\*
  - 11: gb\_hic:\*
  - 12: gb\_est3:\*
  - 13: gb\_est4:\*
  - 14: gb\_est5:\*
  - 15: em\_estfun:\*
  - 16: em\_estom:\*
  - 17: gb\_gss:\*
  - 18: em\_gss\_hum:\*
  - 19: em\_gss\_inv:\*
  - 20: em\_gss\_pln:\*
  - 21: em\_gss\_vit:\*
  - 22: em\_gss\_fun:\*
  - 23: em\_gss\_man:\*
  - 24: em\_gss\_mus:\*
  - 25: em\_gss\_other:\*
  - 26: em\_gss\_pro:\*
  - 27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	140	11.5	615	11 AY108005	AY108005 Zea mays
2	137	11.3	627	10 AW052991	AW052991 614077E08
3	118	9.7	591	12 BG873962	BG873962 MEST44-F0
4	108	8.9	272	10 AW288601	AW288601 618069F05
5	101	8.3	736	17 BH839239	BH839239 LMCRI5000
6	73	6.0	675	10 BE357876	BE357876 DGI_22_A0

7	73	6.0	691	10 BE355111	BE355111 DGI_113_B
8	68	5.6	699	10 BE361874	BE361874 DGI_82_G0
9	68	5.6	721	10 AW922946	AW922946 DGI_47_E0
10	68	5.6	746	10 BE361884	BE361884 DGI_82_H0
11	54	4.4	503	9 AI987561	AI987561 614055F02
12	43	3.5	774	14 BM817061	BM817061 HCOIC03_T
13	39	3.2	741	17 AQ365502	AQ365502 nbx0063N
14	38	3.1	162	10 AW922862	AW922862 DGI_47_E0
15	38	3.1	169	10 BE355228	BE355228 DGI_113_B
16	38	3.1	293	10 BE357711	BE357711 DGI_22_A0
17	37	3.0	257	9 AA633314	AA633314 ng58f07.s
18	37	3.0	452	9 AA834048	AA834048 of25f05.s
19	36	3.0	186	13 BI426846	BI426846 sag08b12.
20	36	3.0	200	9 AL045626	AL045626 DKFZP434P
21	36	3.0	291	9 AA752896	AA752896 97AS0589
22	36	3.0	228	12 BG630772	BG630772 CC-esf1cL
23	36	3.0	366	14 BG523661	BG523661 NISC_n123
24	36	3.0	383	14 RS3048	RS3048 y986b08.s1
25	36	3.0	396	10 BE446576	BE446576 WHE1458_G
26	36	3.0	503	14 BQ196967	BQ196967 NXLV107_F
27	36	3.0	541	9 AI392265	AI392265 NCSFP310T
28	36	3.0	588	17 DR1J17S	AL733665 Danio rer
29	36	3.0	599	9 AI392052	AI392052 NCCSF47
30	36	3.0	600	14 B0412975	B0412975 GA_Ed006
31	36	3.0	719	17 BH838987	BH838987 LMCRI5000
32	36	3.0	818	9 AU100877	AU100877 AU100877
33	35	2.9	259	9 AI920983	AI920983 w018a05.x
34	35	2.9	300	9 AI482375	AI482375 v963b06.x
35	35	2.9	341	10 BE586856	BE586856 WHE0508_C
36	35	2.9	351	12 BG630039	BG630039 CC-esf1cL
37	35	2.9	357	9 AI623587	AI623587 ts37f08.x
38	35	2.9	430	10 AV808136	AV808136 AV808136
39	35	2.9	438	9 AU223340	AU223340 AU223340
40	35	2.9	470	9 AL829048	AL829048 AL829048
41	35	2.9	503	10 AW506165	AW506165 G51497_G1
42	35	2.9	517	9 AL502136	AL502136 AL502136
43	35	2.9	520	13 BM075935	BM075935 MEST362-D
44	35	2.9	533	9 AU088569	AU088569 AU088569
45	35	2.9	540	9 AA897915	AA897915 NCCIG6T7

ALIGNMENTS

RESULT 1  
AY108005  
LOCUS Zea mays PC0076418 mRNA sequence.  
DEFINITION Zea mays  
ACCESSION AY108005  
VERSION AY108005.1 GI:21211083  
KEYWORDS HTC.  
SOURCE Zea mays.  
ORGANISM Zea mays.  
REFERENCE 1 (bases 1 to 615)  
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whiteitt,M.S.,  
TITLE Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.  
JOURNAL Maize Mapping Project/DuPont Consensus Sequences for Design of  
REFERENCE Unpublished (2002)  
AUTHORS Overgo Probes  
TITLE 2 (bases 1 to 615)  
JOURNAL Direct Submission  
AUTHORS Coe,E.C.  
TITLE Submitted (25-APR-2002) Maize Mapping Project, University of  
JOURNAL Missouri, Columbia, MO 65211, USA  
FEATURES  
source  
1..615  
/organism="Zea mays"  
/db\_xref="MaizeDB:63422"  
/db\_xref="taxon:457"  
/clone="PC0076418"

AY108005 615 bp mRNA linear HTC 25-MAY-2002  
Zea mays PC0076418 mRNA sequence.  
AY108005  
AY108005.1 GI:21211083  
HTC.  
Zea mays.  
Zea mays.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Zea.  
1 (bases 1 to 615)  
Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whiteitt,M.S.,  
Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.  
Maize Mapping Project/DuPont Consensus Sequences for Design of  
Overgo Probes  
Unpublished (2002)  
2 (bases 1 to 615)  
Coe,E.C.  
Direct Submission  
Submitted (25-APR-2002) Maize Mapping Project, University of  
Missouri, Columbia, MO 65211, USA  
Location/Qualifiers  
1..615  
/organism="Zea mays"  
/db\_xref="MaizeDB:63422"  
/db\_xref="taxon:457"  
/clone="PC0076418"



Db 212 CGAACCAACCAATTAATGGTGAGATAGTAATTAAGATGAAAGATGGGAATGCGCCCTTCACC 153

QY 973 CAAGCAATGGCCCAAGCAAGTCTCTATGCTACCTTTGAAGA 1011

Db 152 CAAGCAATGGCCCAAGCAAGTCTCTATGCTACCTTTGAAGA 114

RESULT 5  
BH839239/c

LOCUS BH839239 736 bp DNA linear GSS 28-MAY-2002

DEFINITION LMCR150009G06f Zea mays L. Zea mays genomic clone LMCR150009G06f, DNA sequence.

ACCESSION BH839239

VERSION BH839239.1 GI:21237254

KEYWORDS GSS.

SOURCE Zea mays.

ORGANISM Zea mays.

REFERENCE Kim,S.W.; Yu,X.; Lee,M.C.; Main,D. and Wing,R.A.  
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.  
1 (bases 1 to 736)

AUTHORS Kim,S.W.; Yu,X.; Lee,M.C.; Main,D. and Wing,R.A.

TITLE Methyl-filtration genomic sequence from maize

JOURNAL Contact: Wing RA

COMMENT Published (2002)  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: twing@clemson.edu  
Total High Quality bases = 531  
Seq primer: TAATACGACTCATATAGGG  
Class: shotgun  
High quality sequence start: 28  
High quality sequence stop: 713.  
Location/Qualifiers  
1..736  
/organism="Zea mays"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="LMCR150009G06f"  
/clone\_lib="Zea mays L."  
/tissue\_type="Leaf"  
/lab\_host="DH10B"  
/note="vector: pGEM-T easy; Site 1: Mcr BC;  
Methyl-filtration library, Nuclei DNA was completely digested with Mcr BC, size fractionated and transformed to E.Coli.DH10B."

BASE COUNT 164 a 224 c 220 g 123 t 5 others

ORIGIN

Query Match 8.3%; Score 101; DB 17; Length 736;  
Best Local Similarity 100.0%; Pred.No.1.3e-14;  
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 GATAGTGGCGCGCGAGGGGAGGGCGCGGTGCGTACGCACTGCACCCGCGCAATT 328

Db 567 GATAGTGGCGCGCGAGGGGAGGGCGCGGTGCGTACGCACTGCACCCGCGCAATT 508

QY 329 GGGCGGTGGGTGGCGCGCGCGCGGGAATGGAGCAGCGCC 369

Db 507 GGGCGGTGGGTGGCGCGCGCGCGGGAATGGAGCAGCGCC 467

RESULT 6  
BE357876

LOCUS BE357876 675 bp mRNA linear EST 20-JUL-2000

DEFINITION DGI\_22\_A06\_g1\_A002 Dark Grown 1 (DGI) Sorghum bicolor cDNA, mRNA sequence.

ACCESSION BE357876

VERSION BE357876.1 GI:9299433

KEYWORDS EST.

SOURCE sorghum.  
ORGANISM Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.  
REFERENCE 1 (bases 1 to 675)  
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt,L.H.  
TITLE An EST database from Sorghum: dark-grown seedlings  
JOURNAL Unpublished (2000)  
COMMENT Contact: Cordonnier-Pratt MM  
Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmp@pratt@uga.edu  
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.  
Seq primer: PolyTMix  
High quality sequence start: 30  
High quality sequence stop: 638  
POLYA=No.  
FEATURES  
source  
1..675  
/organism="Sorghum bicolor"  
/db\_xref="taxon:4558"  
/clone\_lib="Dark Grown 1 (DG1)"  
/note="Organ: 5-day-old dark-grown seedlings; Vector: Lambda Zap; Site\_1: XhoI; Site\_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."  
BASE COUNT 207 a 121 c 161 g 186 t  
ORIGIN  
Query Match 6.0%; Score 73; DB 10; Length 675;  
Best Local Similarity 100.0%; Pred. No. 3.7e-08;  
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 633 GAGATTAAAGCTTTAAGCGAGAACCGTCGAGAAATATACGAAAGTTGGTGAAGTGG 692  
168 GAGATTAAAGCTTTAAGCGAGAACCGTCGAGAAATATACGAAAGTTGGTGAAGTGG 227  
QY 693 CAAAGCTTTTGGC 705  
DB 228 CAAAGCTTTTGGC 240  
RESULT 7  
BE35111 691 bp mRNA linear EST 20-JUL-2000  
LOCUS DG1.113.B07.g1\_A002 Dark Grown 1 (DG1) Sorghum bicolor cDNA, mRNA  
DEFINITION sequence.  
ACCESSION BE35111  
VERSION BE35111.1 GI:9296311  
KEYWORDS EST  
SOURCE sorghum.  
ORGANISM Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.  
REFERENCE 1 (bases 1 to 691)  
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt,L.H.  
TITLE An EST database from Sorghum: dark-grown seedlings  
JOURNAL Unpublished (2000)  
COMMENT Contact: Cordonnier-Pratt MM  
Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210

Email: mmp@pratt@uga.edu  
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.  
Seq primer: PolyTMix  
High quality sequence start: 23  
High quality sequence stop: 682  
POLYA=No.  
FEATURES  
source  
1..691  
/organism="Sorghum bicolor"  
/db\_xref="taxon:4558"  
/clone\_lib="Dark Grown 1 (DG1)"  
/note="Organ: 5-day-old dark-grown seedlings; Vector: Lambda Zap; Site\_1: XhoI; Site\_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."  
BASE COUNT 212 a 124 c 167 g 188 t  
ORIGIN  
Query Match 6.0%; Score 73; DB 10; Length 691;  
Best Local Similarity 100.0%; Pred. No. 3.6e-08;  
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 633 GAGATTAAAGCTTTAAGCGAGAACCGTCGAGAAATATACGAAAGTTGGTGAAGTGG 692  
161 GAGATTAAAGCTTTAAGCGAGAACCGTCGAGAAATATACGAAAGTTGGTGAAGTGG 220  
QY 693 CAAAGCTTTTGGC 705  
DB 221 CAAAGCTTTTGGC 233  
RESULT 8  
BE361874 699 bp mRNA linear EST 20-JUL-2000  
LOCUS BE361874  
DEFINITION BE361874.g1\_A002 Dark Grown 1 (DG1) Sorghum bicolor cDNA, mRNA  
sequence.  
ACCESSION BE361874  
VERSION BE361874.1 GI:9303431  
KEYWORDS EST  
SOURCE sorghum.  
ORGANISM Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.  
REFERENCE 1 (bases 1 to 699)  
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt,L.H.  
TITLE An EST database from Sorghum: dark-grown seedlings  
JOURNAL Unpublished (2000)  
COMMENT Contact: Cordonnier-Pratt MM  
Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmp@pratt@uga.edu  
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.  
Seq primer: PolyTMix  
High quality sequence start: 64  
High quality sequence stop: 697  
POLYA=No.  
FEATURES  
source  
1..699  
/organism="Sorghum bicolor"  
/db\_xref="taxon:4558"  
/clone\_lib="Dark Grown 1 (DG1)"  
/note="Organ: 5-day-old dark-grown seedlings; Vector: Lambda Zap; Site\_1: XhoI; Site\_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II.

Clones to be sequenced were prepared by mass excision."

BASE COUNT 199 a 121 c 154 g 225 t  
 ORIGIN  
 Query Match 5.6%; Score 68; DB 10; Length 699;  
 Best Local Similarity 100.0%; Pred. No. 5e-07;  
 Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 717 TCATATGCATTGCTAGCTTGATATCTCCATACAGGAGAGATCGTGATGCGGTC 776  
 |||||

DB 32 TCATATGCATTGCTAGCTTGATATCTCCATACAGGAGAGATCGTGATGCGGTC 91  
 |||||

QY 777 TACTTCCA 784  
 |||||

DB 92 TACTTCCA 99  
 |||||

RESULT 9  
 AW922946  
 LOCUS  
 DEFINITION DGI\_47\_E04\_g1\_A002 Dark Grown 1 (DGI) Sorghum bicolor cDNA, mRNA  
 sequence.  
 ACCESSION AW922946  
 VERSION AW922946.1 GI:8088771  
 KEYWORDS EST.  
 SOURCE sorghum.

ORGANISM  
 Sorghum bicolor  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
 clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 721)  
 AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt  
 ,L.H.

TITLE An EST database from Sorghum: dark-grown seedlings

JOURNAL  
 COMMENT Unpublished (2000)  
 Contact: Cordonnier-Pratt MM  
 Laboratory for Genomics and Bioinformatics  
 The University of Georgia, Department of Plant Biology  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 583 0210  
 Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions  
 below Phred quality 16. The threshold for highest quality sequence  
 is 20.  
 Seq primer: PolyTMix  
 High quality sequence start: 23  
 High quality sequence stop: 703  
 POLYA=No.

FEATURES  
 source  
 1..721  
 /organism="Sorghum bicolor"  
 /db\_xref="taxon:4558"  
 /clone\_lib="Dark Grown 1 (DGI)"  
 /note="Organ: 5-day-old dark-grown seedlings; Vector:  
 Lambda Zap; Site 1: XhoI; Site 2: EcoRI; The library was  
 made from poly-A RNA in the cloning vector lambda ZAP II.  
 Clones to be sequenced were prepared by mass excision."  
 202 a 124 c 161 g 234 t

BASE COUNT 202 a 124 c 161 g 234 t  
 ORIGIN  
 Query Match 5.6%; Score 68; DB 10; Length 721;  
 Best Local Similarity 100.0%; Pred. No. 4.9e-07;  
 Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 717 TCATATGCATTGCTAGCTTGATATCTCCATACAGGAGAGATCGTGATGCGGTC 776  
 |||||

DB 41 TCATATGCATTGCTAGCTTGATATCTCCATACAGGAGAGATCGTGATGCGGTC 100  
 |||||

QY 777 TACTTCCA 784  
 |||||

DB 101 TACTTCCA 108  
 |||||

RESULT 10  
 BE361884  
 LOCUS

DEFINITION DGI\_82\_H05\_g1\_A002 Dark Grown 1 (DGI) Sorghum bicolor cDNA, mRNA  
 sequence.  
 ACCESSION BE361884  
 VERSION BE361884.1 GI:9303441  
 KEYWORDS EST.  
 SOURCE sorghum.

ORGANISM  
 Sorghum bicolor  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
 clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 746)  
 AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt  
 ,L.H.

TITLE An EST database from Sorghum: dark-grown seedlings

JOURNAL  
 COMMENT Unpublished (2000)  
 Contact: Cordonnier-Pratt MM  
 Laboratory for Genomics and Bioinformatics  
 The University of Georgia, Department of Plant Biology  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 583 0210  
 Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions  
 below Phred quality 16. The threshold for highest quality sequence  
 is 20.  
 Seq primer: PolyTMix  
 High quality sequence start: 30  
 High quality sequence stop: 719  
 POLYA=No.

FEATURES  
 source  
 1..746  
 /organism="Sorghum bicolor"  
 /db\_xref="taxon:4558"  
 /clone\_lib="Dark Grown 1 (DGI)"  
 /note="Organ: 5-day-old dark-grown seedlings; Vector:  
 Lambda Zap; Site 1: XhoI; Site 2: EcoRI; The library was  
 made from poly-A RNA in the cloning vector lambda ZAP II.  
 Clones to be sequenced were prepared by mass excision."  
 218 a 126 c 166 g 236 t

BASE COUNT 218 a 126 c 166 g 236 t  
 ORIGIN  
 Query Match 5.6%; Score 68; DB 10; Length 746;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-07;  
 Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 717 TCATATGCATTGCTAGCTTGATATCTCCATACAGGAGAGATCGTGATGCGGTC 776  
 |||||

DB 54 TCATATGCATTGCTAGCTTGATATCTCCATACAGGAGAGATCGTGATGCGGTC 113  
 |||||

QY 777 TACTTCCA 784  
 |||||

DB 114 TACTTCCA 121  
 |||||

RESULT 11  
 AI987561/c  
 LOCUS

DEFINITION 614055F02.x1 614 - root cDNA library from Walbot Lab Zea mays cDNA,  
 mRNA sequence.  
 ACCESSION AI987561  
 VERSION AI987561.1 GI:5819355  
 KEYWORDS EST.  
 SOURCE Zea mays.

ORGANISM  
 Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
 clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 503)  
 AUTHORS Walbot,V.

TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University  
JOURNAL Unpublished (1999)  
COMMENT Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 614055 row: F column: 02.  
Location/Qualifiers

FEATURES  
source  
1. .503  
/organism="Zea mays"  
/cultivar="W23"  
/db\_xref="taxon:4577"  
/clone\_lib="614 - root cDNA library from Walbot Lab"  
/tissue\_type="root"  
/dev\_stage="3-4 days old"  
/lab\_host="XLOLR"  
/note="Organ: root; Vector: pBluescriptII SK+; Site: 1; EcORI; Site\_2: XhoI; 3-4 days old root tissue from Walbot Lab (LM)"

BASE COUNT 163 a 118 c 82 g 140 t

ORIGIN  
Query Match 4.4%; Score 54; DB 9; Length 503;  
Best Local Similarity 100.0%; Pred. No. 0.0011;  
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 956 GGAAATGCCCTTACCCAAAGCATGGCCAAAGCAAGTTCTATGCTACTTGGAAGA 1011  
|||||  
DB 466 GGAAATGCCCTTACCCAAAGCATGGCCAAAGCAAGTTCTATGCTACTTGGAAGA 413  
|||||

RESULT 12  
BM617061 774 bp mRNA linear EST 05-MAR-2002  
LOCUS HC01C03.T3.ab1 HC Hordeum vulgare cDNA clone HC01C03.T3.ab1 similar  
DEFINITION to Adenosine Kinase, Nitric-oxide synthase, ATP adenylyltransferase  
to Adenylylsulfate Kinase, Sulfate adenylyltransferase  
sequence.  
ACCESSION BM617061  
VERSION BM617061.1 GI:19153075  
KEYWORDS EST.  
SOURCE Hordeum vulgare.  
ORGANISM Hordeum vulgare.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae  
; Triticeae; Hordeum.  
1 (bases 1 to 774)  
Ozurt,N.Z., Michalowski,C.B., Brazille,S., Borchert,C., Palacio,C.,  
Fredericksen,M.A. and Bohnert,H.J.  
Monitoring large-scale changes in transcript abundance in drought-  
and salt-stressed barley  
unpublished (2002)  
JOURNAL Contact: Mark A. Fredericksen  
COMMENT Plant Biology  
University of Illinois  
1201 W Gregory Dr. Urbana, IL 61801, USA  
Tel: 2172655473  
Email: bohnert@life.uiuc.edu  
Location/Qualifiers

FEATURES  
source  
1. .774  
/organism="Hordeum vulgare"  
/strain="cv tokay"  
/db\_xref="taxon:4513"  
/clone\_lib="HC01C03.T3.ab1"  
/clone\_lib="HC"  
/tissue\_type="root"  
/dev\_stage="3 week old"

BASE COUNT 199 a 151 c 189 g 235 t  
/note="6 and 10 hour drought stress by placing plants on moist paper (75% rel. humidity) in light"

ORIGIN  
Query Match 3.5%; Score 43; DB 14; Length 774;  
Best Local Similarity 100.0%; Pred. No. 0.24;  
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 956 GGGAATGCCCTTACCCAAAGCATGGCCAAAGCAAGTTCTAT 998  
|||||  
DB 284 GGGAATGCCCTTACCCAAAGCATGGCCAAAGCAAGTTCTAT 326  
|||||

RESULT 13  
AQ365502/c 741 bp DNA linear GSS 16-DEC-1999  
LOCUS nbxb0063M14r CUGI Rice BAC library Oryza sativa genomic clone  
DEFINITION nbxb0063M14r, DNA sequence.  
ACCESSION AQ365502  
VERSION AQ365502.2 GI:6585390  
KEYWORDS GSS.  
SOURCE Oryza sativa.  
ORGANISM Oryza sativa.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Erbartoideae; Oryzae; Oryza.  
1 (bases 1 to 741)  
Wing,R.A. and Dean,R.A.  
A BAC End Sequencing Framework to Sequence the Rice Genome  
unpublished (1998)  
On Dec 15, 1999 this sequence version replaced gi:4215156.  
JOURNAL Contact: Wing RA  
COMMENT Clemson University Genomics Institute  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Seg primer: GGAAACACTATGACCATG  
Class: BAC ends  
High quality sequence start: 4  
High quality sequence stop: 386.  
Location/Qualifiers

FEATURES  
source  
1. .741  
/organism="Oryza sativa"  
/strain="Japonica"  
/cultivar="Nipponbare"  
/db\_xref="taxon:4530"  
/clone\_lib="nbxb0063M14r"  
/clone\_lib="CUGI Rice BAC library"  
/tissue\_type="leaf"  
/lab\_host="E. coli DH10B"  
/note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2:  
HindIII; Rice is one of two most popular grains in the  
world. Half of the world population especially those  
inhabiting highly populated areas of the humid tropics  
and subtropics, rely on rice as their primary source of  
carbohydrate. Monocotyledonous rice is a diploid plant  
(2n=24) with a haploid genome equivalent of 431 Mbp  
(Arumuganathan and Earle, 1991). The relatively small  
genome of rice, three times larger than that of  
Arabidopsis, makes it suitable for genomic studies. In  
order to facilitate positional cloning, physical mapping  
and genome sequencing of rice, we have constructed a BAC  
library from Oryza sativa, Nipponbare variety. The  
library contains 36,864 clones with an average insert size  
of 128.5 kb providing 10.9 haploid genome equivalents.  
The deep coverage allows the isolation a particular sequence  
with a probability of 99.9%. Two high density filters,  
each containing 18,432 clones (doubly spotted), represent  
the whole library for colony screening."

BASE COUNT 240 a 148 c 139 g 213 t  
1 others



## ORIGIN

Query Match 3.1%; Score 39; DB 17; Length 741;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 647 AAGCAGAGACCGTGCAGAAATATACGAGGTGGT 685  
 |||||  
 Db 349 AAGCAGAGACCGTGCAGAAATATACGAGGTGGT 311

## RESULT 14

AW922862  
 LOCUS DGI\_47\_E04\_b1\_A002 Dark Grown 1 (DGI) Sorghum bicolor cDNA, mRNA  
 DEFINITION

ACCESSION AW922862  
 VERSION AW922862.1 GI:8088687  
 KEYWORDS EST.  
 SOURCE sorghum.

## ORGANISM

Sorghum bicolor  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
 clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 162)  
 AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt

L.H.  
 TITLE An EST database from Sorghum: dark-grown seedlings

JOURNAL Unpublished (2000)  
 COMMENT Laboratory for Genomics and Bioinformatics  
 The University of Georgia, Department of Plant Biology  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860  
 Fax: 706 583 0210  
 Email: mmpratt@uga.edu  
 Sequences have been trimmed to exclude PolyA, vector and regions  
 below Phred quality 16. The threshold for highest quality sequence  
 is 20.

Seq primer: JEN REV

High quality sequence stop: 161

POLYA=No.

## FEATURES

source

Location/Qualifiers

1..162  
 /organism="Sorghum bicolor"  
 /db\_xref="taxon:4558"  
 /clone\_lib="Dark Grown 1 (DGI)"  
 /note="Organ: 5-day-old dark-grown seedlings; Vector:  
 Lambda Zap; Site\_1: XhoI; Site\_2: EcoRI; The library was  
 made from poly-A RNA in the cloning vector lambda ZAP II.  
 Clones to be sequenced were prepared by mass excision."  
 BASE COUNT 20 a 67 c 53 g 22 t

## ORIGIN

Query Match 3.1%; Score 38; DB 10; Length 162;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 206 CCGCGCTCGTGCATGCCTCACTCCCGTTCCTC 243  
 |||||  
 Db 67 CCGCGCTCGTGCATGCCTCACTCCCGTTCCTC 104

## RESULT 15

BE355228  
 LOCUS DGI\_113\_E07\_b1\_A002 Dark Grown 1 (DGI) Sorghum bicolor cDNA, mRNA  
 DEFINITION

ACCESSION BE355228  
 VERSION BE355228.1 GI:9296232  
 KEYWORDS EST.  
 SOURCE sorghum.

## ORGANISM

Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
 clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 169)  
 AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt

L.H.  
 TITLE An EST database from Sorghum: dark-grown seedlings

JOURNAL Unpublished (2000)  
 COMMENT Contact: Cordonnier-Pratt MM  
 Laboratory for Genomics and Bioinformatics  
 The University of Georgia, Department of Plant Biology  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860  
 Fax: 706 583 0210  
 Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions  
 below Phred quality 16. The threshold for highest quality sequence  
 is 20.

Seq primer: JEN REV

High quality sequence stop: 168

POLYA=No.

## FEATURES

source

Location/Qualifiers

1..169  
 /organism="Sorghum bicolor"  
 /db\_xref="taxon:4558"  
 /clone\_lib="Dark Grown 1 (DGI)"  
 /note="Organ: 5-day-old dark-grown seedlings; Vector:  
 Lambda Zap; Site\_1: XhoI; Site\_2: EcoRI; The library was  
 made from poly-A RNA in the cloning vector lambda ZAP II.  
 Clones to be sequenced were prepared by mass excision."  
 BASE COUNT 21 a 71 c 55 g 22 t

## ORIGIN

Query Match 3.1%; Score 38; DB 10; Length 169;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 206 CCGCGCTCGTGCATGCCTCACTCCCGTTCCTC 243  
 |||||  
 Db 79 CCGCGCTCGTGCATGCCTCACTCCCGTTCCTC 116

Search completed: December 17, 2002, 19:23:00

Job time : 2414 secs

